Journal of
Clinical Psychology

JULY 1957
Volume XIII Number 3

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Editorial Office: 5 Pearl Street
Brandon, Vermont

JERRY W. CARTER, JR.
U. S. Public Health Service

HOWARD F. HUNT
University of Chicago

WILLIAM A. HUNT
Northwestern University

ELAINE F. KINDER
New York State Department of Mental Hygiene

MAURICE LORR
Veterans Administration Washington, D. C.

ROBERT I. WATSON
Northwestern University

ELEANOR C. THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street · · Brandon, Vermont

COMPANION VOLUMES

PRINCIPLES OF PSYCHOLOGICAL EXAMINING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $7.50

PRINCIPLES OF PERSONALITY COUNSELING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $6.00

SYSTEMATIC: The first comprehensive system of diagnosis and psychotherapy based solidly on basic psychological science.

THEORETICAL: Presenting a comprehensive theory of the nature of personality integration with special emphasis on the differential diagnosis and modification of the primary etiologic factors which actually organize socially significant behaviors.

ECLECTIC: All pertinent theories and facts in basic science psychology and clinical psychological science are related in a systematic applied integrative psychology.

COMPLETE: For all teaching and reference purposes.

ADVANCED: Both of these books present a new orientation to diagnosis, counseling and psychotherapy which is more than just a rehash of the literature. Principles of Psychological Examining is more than a textbook of psychometrics since tests and measures are presented only as they relate to more fundamental problems of psychological diagnosis. Similarly, Principles of Personality Counseling presents thorough discussions of the indications and contraindications for applying all known methods of psychotherapy after a primary diagnosis of factors requiring modification has been accomplished.

ORDER NOW

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

JOURNAL OF CLINICAL PSYCHOLOGY

VOL. XIII JULY 1957 No. 3

CONTENTS

GENERAL REVIEWS

Research in Clinical Psychology: Factor Analytic Contributions. W. GRANT DAHLSTROM - 211

The Conceptual and Test Distinction of Neuroticism and Anxiety. RAYMOND B. CATTELL - 221

Critique of Recent Developments in Personality Counseling Theory. FREDERICK C. THORNE - 234

DIAGNOSTIC PROCESSES: GENERAL

The Construct Validity of the Sixteen Personality Factors Test. SAMUEL KARSON and KENNETH BRYNER POOL - - - - - - - - - - - - - - - 245

Confirmation of Nine Psychotic Symptom Patterns. MAURICE LORR, JAMES P. O'CONNOR and JOHN W. STAFFORD - - - - - - - - - - - - - - 252

The Loyola Language Study. VINCENT V. HERR - - - - - - - - - - 258

Measuring the Impairment of the Abstracting Function with the Proverbs Test. CLYDE M. ELMORE and DONALD R. GORHAM - - - - - - - - - 263

Evaluation of Intellectual Potential in Hemiplegic Individuals. FRANKLIN C. SHONTZ - 267

The Prediction of Adjustment over a Four Year Interval. HARRY BEILIN - 270

DIAGNOSTIC PROCESSES: RORSCHACH STUDIES

Factor Analysis of Rorschach Scores of Schizophrenics. BERNARD A. STOTSKY - 275

Effects of Different Methods of Structuring the Rorschach Inquiry upon Determinant Scores. STANLEY W. KLINGENSMITH - - - - - - - - - 279

Psychometric Scales for the Rorschach Popular Response. DOROTHY PARK GRIFFIN - 283

DIAGNOSTIC PROCESSES: SELF AND PEER PERSONALITY RATINGS

Self Congruence, by Sex and Between the Sexes, in a "Normal" Population. JOHN G. MARTIRE and ROBERT H. HORNBERGER - - - - - - - - - - 288

Similarity of Personality in the Sociometric Preferences of Mental Patients. WILLIAM G. SHIPMAN - 292

Self and Peer Personality Ratings of Psychotherapists. ELI A. RUBINSTEIN and MAURICE LORR - 295

CONTENTS—Continued

The Scatter Pattern of a Southern Negro Group on the Wechsler-Bellevue Intelligence Scale. JULIAN C. DAVIS - - - - - - - - - - - - - - - - - 298

The Utilization of a Psychologist in Ward Administration. WARREN W. WEBB and HIRAM L. GORDON - 301

Evaluation of Modified Administration of the California Psychological Inventory. LAWRENCE A. BENNETT and ALVIN RUDOFF - - - - - - - - 303

Seasonal Changes in Admission Rates of a State Hospital. PAUL A. HAUCK - 305

Personality Changes Following Transorbital Lobotomy. NANCY B. ALGEO and MAXWELL S. PULLEN - - - - - - - - - - - - - - - - - - 308

Editorial Opinion - 310

Book Notices - 311

GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Special sections in each issue will be devoted to news, comments, and letters to the editors on matters pertinent to the profession. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

ORIGINAL ARTICLES are published only with the understanding that they are contributed exclusively to the *Journal of Clinical Psychology*. The publishers are not responsible for statements made or opinions expressed by contributors in articles printed in these columns. Copyrights cover publication in the *Journal of Clinical Psychology* and articles may not be reproduced without permission of the publishers.

MANUSCRIPTS must be typewritten, double-spaced with good margins, on one side of the paper only. The publishers will not be responsible for manuscripts lost in the mails or through unavoidable accident. Retain a carbon copy after sending the original to the editor. It is requested that all articles be as brief as consistent with clarity. All manuscripts are subject to editorial modification.

THE STYLE of all manuscripts including bibliographies must be in accordance with the instructions given in: Anderson, J. E., and Valentine, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 41, 345-376. References should be arranged alphabetically according to the directions given above. Full address of the author should appear somewhere in the article, preferably at the end.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscribers' expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $8.50; Canadian subscriptions $9.00; foreign subscriptions $9.50 including postage; single copies $2.50. Entered as second class matter January 15, 1945, at the post office at Burlington, Vermont, under the Act of March 3, 1879. Reentered at the post office at Brandon, Vermont. Additional entry at the post-office at Rutland, Vermont. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont.

Copyright, 1957, Frederick C. Thorne, M. D.

RESEARCH IN CLINICAL PSYCHOLOGY: FACTOR ANALYTIC CONTRIBUTIONS

W. GRANT DAHLSTROM

University of North Carolina

General Theory

Factor analytic techniques are having a significant and growing impact on research in clinical psychology, both from the number of publications in recent years in which factor analytic methods have been employed, and from the contribution factor analytic models have made to the conceptual framework in which ability, temperament and personality are viewed. Clinicians cannot comfortably remain in ignorance about the conceptual, methodological or logical ramifications of factor analysis and still try to keep up with the substantive contributions stemming from the application of these techniques. In this review, an attempt will be made to epitomize the factor analytic approach, describe some of its limitations and survey some of the major theoretical issues being subjected to factor analytic study now.

An excellent non-mathematical treatment of the logic of standard factor analysis has been provided by Hall and Lindsey[28]. Only a few major points can be made here. Variation is the basic concern of any science, the origin of explanatory schemas and the source of criteria for predictions. Two sources contribute variation for the study of personality, inter-individual and intra-individual differences. The assumption is made that differences between various persons reveal the operation of important variables which are pertinent in forming the personality characteristics of any one person, even though such characteristics may not be constant. Differences arising from changes in the reactions of an individual over a period of time similarly disclose factors affecting less permanent personological states. Personality description is made in terms of these comparisons; scientific study is directed to the search for the relevant determiners of these variations. Factor analysis proceeds on the assumption that this scientific study is facilitated by an intermediate step of seeking descriptive economy in a small set of variables labelled variously traits, abilities, or factors. These few variables are not offered necessarily as explanations or as bases of prediction, but as useful alternatives to the vast array of possible comparisons arising from combinations of various specific values of these factors. Thurstone[61] in a general characterization of this step described it as an attempt to reduce significantly the number of degrees of freedom in the construct set from the number in the set of original observations.

The classic scientific method calls for the direct manipulation of one variable, the independent one, and the close scrutiny of its effect on another, the dependent variable, while all else is held as constant as possible. In factor analysis, too, the emphasis is on covariation. The operation of a variable is to be observed, not by directly manipulating it, but by allowing it to vary and scrutinizing the correlations it generates amongst a large set of observations. Some writers say that factor analysis enables the investigator to examine simultaneously the dependencies among a large set of variables, and through statistical rather than manipulative control, run several "experiments" at the same time. Thus, factor analysis is offered as the method of choice when the investigator is in considerable ignorance about the area under study; ignorant, that is, about the variables, the relationships, and the requisite manipulations. It is also offered as suitable when the requisite manipulations are prohibited, as in the areas of mental illness and intellective abilities.

In this context of sheer exploration, several assumptions are advanced as first approximations only (cf. Guttman[27]). Consideration is given only to variables which are linearly related in the observational set. Also, many factor analysts assume orthogonal independence amongst the factors in the construct set. In addi-

tion, the usual specification equation for any one individual's performance on any one observational variable (test) is based on a simple additive combination of products of factor weights of task and person. These are tentative decisions to be used for theory generation and to be changed when the evidence forces some alternative mathematical form. Working assumptions such as these hypotheses provide are necessary, but some additional requirements must also be met if the investigator is to be able to proceed beyond these first approximations. There must be a way of knowing when one is dealing with the same source of variance in several different investigations, so one may determine whether some non-linear relationship is operative in the construct set. There must exist some sampling model to be used as a guide as to when to reject an assumption of linearity or simple additivity. Most importantly, there must be freedom from implicit assumptions in procedure that force the data into the prescribed form. In this last respect, factor analytic methods carry over from the ability domain many hidden assumptions about the dimensions, their reality status and their interrelationships.

The Domain of Factorial Studies

It is regrettable that factor analysis was not advanced as a special instance of analysis of variance. As these general procedures are now described, no one would offer as reasonable investigative behavior the blind partitioning of the variances obtained from sets of observations. The guiding principle in the partitioning of the experimental data in analysis of variance stems from psychological theory and from the observations or manipulations of the independent variable that are generated by the theory. The statistical considerations are neutral in regard to the theory involved. Factor analysis provides a way of partitioning the covariational data, but the lines of segmentation should come from a careful articulation of theory and observation. No procedure, even rotation to simple structure, should be exercised without consideration of the theoretical implications of the set of observations. Many writers have implied that factor analysis is a useful procedure only in the stage where the investigator has no theory and that it loses its usefulness when the investigator becomes an experimental manipulator. This seems wrong in two respects: a great deal needs to be known to use factor analysis appropriately in an area, and the more known the better, while there is no reason to believe that the adroit use of factor analysis cannot help clarify problems in areas where direct manipulation has long been employed with great power (cf. Lewis, McAllister, and Bechtoldt[35, 36]).

The first necessity for detailed knowledge in the application of factor analysis to a problem area is in the definition of the domain itself. The problems involved here in terms of the arbitrariness of the decisions and their effect on the subsequent findings can be illustrated by a familiar analogy.

> Let us consider a psychologist using the well known Object Sorting Test of Goldstein and Scheerer[38]. In laying out the materials in front of the subject, the investigator may choose, say, several pieces of cutlery, a pencil, a letter opener, an ashtray, a pipe, and two cigars, one of tobacco and the other of candy. With this particular array of objects, the sorter might, if he were taking a variety of abstract attitudes, place the objects in classes which involve cutting edges, objects which pertain to smoking, objects made of metal, etc. He also might form a class of objects which are placed in the mouth. It is clear that the particular concepts which the subject is most likely to come up with are a function of the particular objects presented. We are more likely to get the subject to sort the objects into classes relating to smoking if, in addition to the cigar, pipe, and ashtray, we add the matchbox and matches. In this case, the candy cigar might not be placed in this class. However, if we add some objects not generally employed in the test like a candy bar and a penny sucker we would be much more likely to get the subject to sort out a class of confections. In fact, we could change the behavior of the subject rather drastically if we added not just one candy bar but a variety of candy bars, differing in presence of chocolate, nuts, caramel or even went further and added a variety of forms of confections sampling widely from a complete candy store. Under these circumstances the subject may well form a variety of groups of the confections, meanwhile placing all of the non-candy items into one class.

The various factor analytic methods advanced, including the centroid method of Thurstone with rotation to simple structure, prove to be sensitive to the "elements"

which are placed in the matrix as is the behavior of the hypothetical test subject above. If in our ignorance of a psychological domain we include only one test which reflects some particular source of variance, x, then this source of variance will not appear in the common factor matrix. In order for the factor x to appear we would have to include several measures of this source of variance; from the resulting intercorrelations amongst these several x-related measures, we would get common factor space reflecting this variance. In the very early stages of investigation of an area, however, the psychologist is apt to omit a particular line of study from his matrix, or represent it inadequately, so that the common factor space does not include this variable. Similarly, when additional observation variables are added to the matrix, the relationships amongst the factors are altered. Thurstone[51] has advanced the simple structure solution to provide such metric invariance. It was also intended to provide configural invariance in the factor solutions when the same observations are made on different populations of subjects. Recent publications have raised doubts that invariance in simple structural solutions can be assured in either of these situations. As one example, Fleishman and Hempel[23], by adding tasks to a psychomotor battery, split a previous factor, Psychomotor Coordination, into two factors, one emphasizing fine, highly controlled adjustments, the other involving gross coordination of more extensive movements. Addition of tasks to a battery not only may extend the common factor space by bringing into consideration variance previously represented only on one task, but may accentuate differences in variance hitherto assigned to one source.

A similar point can be made from the study of Kettner *et al* [32] in which Thurstone's simple structure was not employed but in which a test, Problem Solving, failed to show its familiar relationship with a general reasoning dimension, presumably because the task was too easy for the selected group of Coast Guard cadets. In both of these studies cited above, the departures from expectation had to be interpreted qualitatively in the light of past experience with the materials and not on the basis of clear-cut evidence of greater than chance variation in factor loadings. The solution to this problem would appear to lie in the judicious, theory-based use of reference tests about which a great deal is already known, together with appropriate sampling models along the lines suggested by Lawley[33] or Rao[42].

A further point should be made about this problem. Formal means for comparing factor analytic solutions from separate studies have not been developed in detail. Vandenberg[53], for example, studied the effects on a group of abilities and their intercorrelations when translated into another language. He gave thirty-five tests to bilingual Chinese students and compared the factor structures obtained in English and Chinese versions. He based his comparisons on Tucker ϕ_r, an index of similarity between factor structures. Ahmavaara[1, 2] also has developed a method he calls transformation analysis for direct comparison of factor solutions from separate studies. Unfortunately, these indices must be interpreted qualitatively as yet since they lack sampling specifications.

In this problem of the definition of the domain, it has usually been assumed that as the knowledge of an area accumulated, the domain would gradually be redefined and appropriately clarified. This hope is certainly a reasonable one, but the usual factor analysis does not provide a seminal basis for such restructuring of the area. The difficulty is that in the factor plots, the dimensions do not differ in ways that would suggest the levels of conceptualization required to fit the observations into a broader theoretical framework. In the ability area, some innovations have been introduced that stemmed mainly from factor analytic surveys, but a great deal of credit historically should go to the non-factor analytic work of Binet, of Terman, and of Pintner, for example, in clarifying the proper area of study in the intellective domain. The necessity for empirical ties with non-test criteria has been overlooked in much factor analytic work on personality tests, to the detriment of many carefully constructed instruments. This requirement is explicitly recognized by Eysenck[20] in his criterion analysis step prior to further factor analytic work.

The Interpretation of Factor Plots

The lack of inherent "meaning" in the factor plots is not necessarily a limitation. No one is worried that the error term in an analysis of variance does not provide in itself clues as to the sources of uncontrolled variation. The point is that workers have advanced factor analysis techniques and results as somehow theory-generating. Currently there is a great deal of concern about the contribution of the examiner to variation in personality test performance. The examiners may well show up in a factor analysis as a source of variation amongst a set of variables, but this "dimension" would not be in any way different than the effects of a certain response set or instructional difficulty, when represented in the factor plots.

Although some writers advance factor analytic results as mere descriptive economy, many write and act upon implicit or explicit existential hypotheses concerning the factors. For them the factors are there, the investigator has merely to find them. A careful treatment of these dangers of reification in psychological constructs can be found in Bergmann[4], but a few comments are pertinent here in the context of factor analytic work. The issue is subtle but important since one method of factor analysis or another, or one method of rotation or another, is advanced because it will provide the identification of the "real" factors.

A form of this issue appears in the work of Rosenthal[44] on the problem of anxiety. For her purposes she employed 54 variables alleged to measure anxiety in a long list of studies in the area. She posed her problem as an attempt to determine "whether one concept is referred to in these studies" or not. She included questionnaire data, physiological measures, conditioning and learning variables. She intercorrelated these measures on 70 students and performed a centroid factor analysis with a rotation to simple structure. In one sense, she could have answered her question directly by inspection of the intercorrelational matrix. If all the variables did not fall within the positive manifold, that is, if all the variables were not significantly and positively correlated, she would have had an answer to her problem. She found that they did not. Secondly, she just as legitimately could have employed a centroid factor analysis without rotation. Examination of the first factor loadings, the way that each of the measures was related to the general mean of the test vectors, would have demonstrated the usefulness of a single construct. Since she went to some pains to be exhaustive in regard to the important measures advanced for anxiety, she also could have justified the use of Hotelling's[30] principal component method (placing values of unity in the diagonals in the matrix) and the values of each of the measures of the first principal component would have been equally informative. Since, however, she did extract 11 factors and rotated these to an oblique structure, the question under study should probably have been addressed to a second order factor analysis of the loadings on these 11 factors. Granted that particular errors or systematic non-anxiety variance enters each or all of the measures in her matrix, her attention really should be directed toward the role of one central second order factor amongst her group factors. If such a single source of variance was found this would be a legitimate definition of anxiety for her purposes. It is unfortunately impossible to tell from her excellent abstract how much attention she paid to these considerations in her full report.

These implications concerning the reality status of the factors also appear again and again in the modifications of factor analytic methods involving cluster search or type-finding techniques (cf. Sakoda[45]). One of the recent modifications receiving a great deal of attention is the D function of Cronbach and Gleser[14] which expresses the similarity between test patterns. As Lykken[37] comments about this statistic, "Only by remote coincidence would a measure of the dimension of geometrical distance serve also as a metric for a psychological dimension, and a suitable index of similarity for (some) one attribute could not, by definition, measure similarity in (some) other attribute if the two are truly independent." Many factor studies seem to be bound by servile conformity to some particular mechanical solution in the false belief that in some magical way the method will uncover the hidden, basic truths.

Factor Analysis and Psychological Theory

The full and appropriate application of factor analytic techniques is further hindered by its historical relationship to a sterile, clumsy trait theory. Work in the area has not led as yet to anything resembling, in Bergmann's [6] terms, process laws of psychological variables, nor to any serious consideration of the combinatorial problems which await empirical investigation. Much of this work has suffered from the inherent difficulties of what Bergmann [5] and Spence [46] have termed R-R formulations which better serve taxonomic and classificatory purposes than the aims of generating theory. The taxonomic problem may be resolved by a multiple correlation and its related beta weights, or a linear discriminant function, but the simple addition of variables according to some geometric model or vector schema does not meet the requirements of behavior theory at the current level of sophistication. The construct of anxiety, mentioned above in the study of Miss Rosenthal, serves as a good example of this point. Factor analytic methods may well furnish evidence and information that would lead to some construct like anxiety as a source of individual differences. In fact, the description Miss Rosenthal gives of her factor 1 is very close to Taylor's Manifest Anxiety Scale or Welsh's Judged Anxiety Scale and his first Factor Scale, A. But there is little basis within the traditional use of factor analysis to expect that the additional specifications about its conditionability (and extinguishability), its drive properties, its cue-producing function, its role in the reinforcing process, its enhancement of behavioral efficiency under some specific conditions and its inhibitory or disruptive effects on on-going behavior under slightly altered conditions could easily be obtained merely from continued factor analytic exploration of the domain. These important considerations have been assigned to anxiety as a consequence of the introduction of this construct into a framework of behavior theory according to clearly specified rules of scientific theory building. Many of these interrelationships are not monotonic, few of them are simply additive, many of them are highly dependent upon momentary quantitative values of both the anxiety variable and other constructs in the system. The great importance these problems of personality organization have in current research (cf. Thorndike [50]) makes it all the more deplorable that factor analytic work continues without properties with a sophisticated psychological theory (cf. Franks [25]).

Factor Studies of W-B Patterning

One of the most widely studied problems in current clinical research has been the problem of Wechsler-Bellevue patterning. Some of the key issues in this area could have been cast in a form in which factor analytic methods might have been usefully applied. One form, spelled out in most detail by Jastak [31] in his concentric method of test analysis, was that all of the forms of intellective behavior sampled (on the subtests of the W-B, for example) reflect one basic ability and therefore a person with a given ability should be expected to have the same standing on any of the measures. When he does not, the deficiencies appearing on some of the subtests reveal areas in which he has not been developed to full capacity. The level one can expect the person to develop is set at the level of the highest performance in the battery, *i.e.*, Jastak's *altitude score*. Several problems are involved in the interpretation of test profiles in this way (cf. Brown and Bryan [7]), including the fact that even the highest value could itself also be "depressed" to an unknown degree and the fact that this whole approach ignores the problem of errors of measurement, which are undoubtedly different for each of the subtests. But the crucial question is whether each of the subtests can be considered a homogeneous measure of two sources of variance: a source common to all of the subtests and a source specific to the subtest itself. If not, the concentric model cannot be appropriately applied, and the inferences drawn from the profile must be considered erroneous.

The analysis of Davis [17] was based on a matrix which included not only the W-B subtests but also reference tests from Air Force studies and other sources that were well known in terms of their relationships within the intellective domain. He

used these reference tests to help determine his final rotational solution and found that in these terms, all of the subtests of the W-B were reflecting multiple sources of variance for this junior high school population. The vocabulary, arithmetic, and picture arrangement subtests were the least complex. The comprehension subtest was found to be very complex in his solution but even so the amount of its variation reflected in this battery was small. Each of the sources of variance from this common factor space in the comprehension subtest had a higher representation on some other subtest in the W-B. From data like those provided by Davis, considerable doubt exists that the W-B meets the assumptions listed above for the "single-factor" interpretation of the profile.

Another form of the patterning research on the W-B is that all of the variance in the subtests that is not attributable to the intelligence dimension for which the W-B was constructed is a function of different non-intellective sources of variance that are important in evaluating the subjects' personality status. This question, too, is easily submitted to factor analytic scrutiny. Frank [24] approached this question by means of a p-technique analysis in which persons are correlated, rather than the standard r-technique in which tests are correlated. Burt [8] has demonstrated the algebraic relationships existing between these two procedures when they are based on the same tests and subjects. Stephenson [48] has extended the procedures based on the correlation of persons beyond this by using samples in which the interchangeability mentioned above is no longer possible. The procedure of correlating persons involves the reduction of the separate measurements involved in the correlations to standard scores, which has been done in the weighted scores of the W-B. Frank took the additional step needed when the number of observations is small of using rank-order correlations for his original matrix. His subjects came from the work of Rapaport *et al.* [43] and included 60 neuropsychiatric patients differing in sex, age, intelligence, and psychiatric diagnosis. He found two dimensions, which in p-technique appear as types of persons. These groupings did not show any homogeneity in respect to age, sex, or diagnosis. Rather, the sources of variance were most economically ascribed to performance and verbal abilities. In his brief report, insufficient data were provided to enable the reader to judge the evidence he used for stopping at two dimensions but if he appropriately followed Thurstone in the application of the centroid method of factor analysis, little additional variance remained in the intercorrelational matrix after extracting these two factors. Inspection of the magnitudes of the loadings in his factor table would seem to support this. This evidence is consistent with the findings of Wittenborn and Holzberg [55] and should be sobering to those who are prone to generate elaborate speculations on the basis of highs and lows in ability profiles on psychiatric cases.

A study of differences in the interrelationships existing among various intellective measures in groups of children of different socio-economic levels which was reported by Mitchell [39] has bearing on the very important question of non-intellective biases in our diagnostic instruments. He employed a direct measure of S-E status for grouping his subjects that were initially drawn from schools which were known to differ in their sampling characteristics on this dimension. He employed an array of well-constructed reference tests from the Psychometric Laboratory of Thurstone and the short-form version of the California Test of Mental Maturity. He also included two of the kinds of tests most criticized by Davis and Hess [16] as biased in favor of middle class groups, the Otis and the Henmon-Nelson. One of his most important findings concerned the higher interrelationships amongst the usual ability variables in the lower S-E groups, suggesting less differentiation amongst the requisite skills. For the low S-E group the verbal features of the tasks posed the problem, either in the instructions themselves or in the wording of the item, so that these tests were just variations on reading and verbal comprehension tasks. This difference in structure was not eliminated by taking sub-samples from his two groups that were matched for scores on the Otis but preserved the original S-E differences in the total groups. The question of the possible influence on these

structures of differences in variances on the original measurements was also explored and dismissed. The influence of this lack of ability differentiation on some of our standard diagnostic instruments should be investigated.

Factor Analytic Studies of Psychiatric Diagnostic Systems

The importance of the psychiatric diagnostic system as a source of criteria in test construction, pattern analysis and evaluations of therapeutic techniques makes the clarification of this domain by factor analytic methods of particular interest. A great deal of the work along this line from Moore [40] on has been based upon symptoms or characteristics reported by observers or therapists. The problems of lack of reproduceability of these original observations and the lack of experimental independence amongst the set of ratings or judgments made by a single observer on a single assessment sample have been particularly plaguing. Fiske [22] has pointed out very cogently that the factor dimensions may reflect to an unknown extent the halos and stereotypes of the raters and have little reference to the behavior patterns of the patients under study. A recent exchange of comments between Conger *et al.* [12, 13] and Stephenson [47] has served to point up the way rater variables have confounded the work of Beck on schizophrenic types [3]. Since Beck included correlations in the matrices to be analyzed that were importantly a function of the raters themselves, the particular groupings of patients he formed amongst the schizophrenics he studied must be examined carefully and subjected to independent confirmation on new groups of cases and observers.

Another integrated series of projects in the psychiatric domain that has lead to the espousal of three main dimensions of variation is that of Eysenck and his coworkers [18, 19, 20, 21]. They have considered these taxonomic dimensions to be extraversion-introversion, neuroticism and psychoticism. Workers in the psychiatric area are generally surprised that Eysenck is willing to settle for so few basic dimensions in a domain in which most observers have felt the need for many more ways of describing the observed differences. The way that this economy is achieved can be seen from a few examples from Eysenck's analyses. For him, "neurosis" is mainly dyspepsia and hypochondriasis [19, p. 34-35], while hysteria is comprised largely of "extraversion." Psychasthenia (obsessional) is mostly "introversion"; little neurotic loadings appear for either of these latter two diagnostic designations. Combinations would be used to describe depression, anxiety, and the like.

There has been some question, however, whether Eysenck's economical system was more a function of the limitations of the material included in his matrices than any revolutionary way of recasting the domain of psychiatric nosology. Recent evidence on this point comes from a study from the Maudsley group. Trouton and Maxwell [52] studied a large group of male psychiatric patients and obtained data from a checklist of observations routinely recorded on all cases admitted to the unit. They selected 45 variables which would cover the domain as specified by Cattell [9] and intercorrelated them by tetrachoric methods after assuring themselves that each item occurred in at least ten percent of the group. Six factors were extracted by Thurstone's centroid method before the residuals approached the chance level. The authors identified two of these sources of variance with Eysenck's neuroticism and psychoticism, but found such important considerations in psychiatric status as thought disorder, depression, social withdrawal, and possibly hypochondriasis and obsessionality appearing in important degrees on the four other dimensions. After a careful consideration of some of the sources of bias in this method, such as semantic relations amongst synonyms and antonyms, the fluctuations in behavior over an entire period on the psychiatric service, and arbitrary separation of similar features because of psychiatric formalities in mental status description, the authors concluded that neither the usual psychiatric diagnostic system nor the first order orthogonal factor structure provided a clear framework for the classification of the observations. They expressed some hope, however, that such a schema could come from a second-order analysis amongst the rotated first order factors.

Applications to Personality Tests

As Cattell[10] indicates, there are a restricted number of personality instruments in existence that have had factor analytic scrutiny of either the homogeneity or the independence of the component scales. It is equally noteworthy that such instruments that are available have proven to be of limited utility, impractical and relatively unpopular. To this reviewer, the difficulty seems to lie in the omission of the step mentioned earlier as "criterion analysis". Although the need for external criteria has been recognized for a long time by such empiricists as Strong for the *SVIP* and Hathaway for the MMPI, the development of tests by factor analytic methods has not proceeded along this line. It is a little disturbing to find that Cattell, in describing the reworking of his elaborate 16 P. F. questionnaire, employs the highly honorific terms, "cross-validation"[49] and "construct validity",[15] in radically altered meaning. Working with different samples of highly selected groups of college students or Air Force men, Cattell claims to have cross-validated the items in his various scales by demonstrating that the items fall on the same dimensions in the factor plots of the new samples that they fell on in the original sample structure. The structures in the new samples were heavily determined by these very items, however, which makes this invariance less surprising. The more disturbing feature of this circular argument is that none of this material furnishes any evidence on the pertinence of the interpretive labels applied to the scales. Interpretation must be made within a body of psychological knowledge. Validity means that one can reasonably make the logical substitution implied in the scale title. Cross-validation means that one has independent evidence about the justifiability of these inferences, free of the specific sampling characteristics of the original group on which the instrument was developed and the local idiosyncrasies of the non-test criteria. For Cattell's tests there is neither freedom, since the criteria are not independent of the test or of the original developmental details. Construct validity, as employed by Meehl and Cronbach[15], implies that the test scores relate to non-test variables in a way required of them within the nomological net of psychological theory if the scale is validly related to the referent construct within that theory. For Cattell, however, this use is abandoned and he has equated construct validity with internal consistency methods[10, p. 205]. For this well-worn notion, he seeks to substitute the new term "intensification".

Factor Studies of the MMPI

There is a large literature now available on the attempts to re-evaluate the empirically derived MMPI by means of factor analytic techniques. For a great deal of this work the point mentioned previously about the importance of careful definition of the domain must be re-emphasized. For some of the work there is an explicit interest in the variance common to the clinical scales. Welsh[54], for example, was trying to clarify the construct of "general maladjustment" that would lead to positive intercorrelations amongst the separate scales. When attention is directed to the sources of variance reflected in some one scale, or in the MMPI item pool, then non-MMPI variables must be included in the matrices. An example of this problem is the study of O'Connor *et al.*[41] on the MMPI depression scale. They computed the intercorrelations of 49 of the 60 D scale items based on the replies of 300 V. A. neuropsychiatric cases, factored the resulting matrix and reported five sources of variance amongst the scale responses. The authors conclude, "The results indicate the multidimensionality of the D scale and call into question the practice of attributing to the scale a simple unitary significance." This conclusion is either trivial or premature, depending on the emphasis given to it. A mountain of evidence is now available that the multiphasic nature of personality inventories lies at the item level. Every item in the MMPI can be and is interpreted differentially, by different subjects and by the same subject at different times. It is now trivial therefore that evidence should be offered that the items do not behave as a Guttman-type scale when grouped in the D scale of the MMPI. However, there is still the possibility that

the psychological construct, depression, as employed in the psychiatric nomological net would demonstrate the same multidimensional properties assigned to the D scale by these investigators. Although it does not seem likely that this would be the case, one cannot decide the question of the D scale's "simple unitary significance" without including in the factor analysis some non-MMPI variables related to the construct, depression. In a similar way, studies purporting to show that the MMPI is only two-dimensional, or three-dimensional, and not a nine-scale clinical instrument are equally misleading. The dimensionality of the MMPI pool is literally infinite: any reasonably intelligent respondent could go on endlessly forming and reforming classes into which MMPI items could be assembled. One of the most promising subsets of these classes now under investigation is formed by the configural scoring methods of Meehl [38] and of Horst [29], but the dimensions to be "discovered" within this subset alone are probably limitless, too.

Skillful use of factor analytic methods in development of personality test instruments, in which the methods do not seem to dominate the psychological theory, is being shown in such different areas as the TAT-related need achievement measure by Clark and McClelland [11], the role repertory by Levy and Dugan [34], and the factor scales by Welsh [64].

Summary

In spite of the forbidding mathematical hurdles and computational labors involved, factor analytic techniques are currently very popular and are having a significant impact upon research in clinical psychology. Seeking clarification of areas in which little is known, or desiring some statistical control over variables not directly under manipulative control, investigators have adopted factor analysis as a way out of these difficulties. Unfortunately they often have been heedless of the limitations of the techniques, ignorant of the assumptions underlying the methods, and uncritical of the burden of implicit psychological connotations carried by the factor analytic rationale. The nature of some of these hypotheses concerning the reality status, distributional forms and interrelational organization of factors and the implications they may have on the design, execution and interpretation of investigations were examined in some detail. The need for closer articulation with psychological theory in the application of factor analysis was emphasized and the need to free factor analytic methods from some of the unwarranted theoretical and psychological connotations they now carry was described.

Bibliography

1. Ahmavaara, Y. Transformation analysis of factorial data. *Ann. Acad. Scient. Fenn.*, 1954, 88, 1-150.
2. Ahmavaara, Y. On the unified factor theory of mind. *Ann. Acad. Scient. Fenn.*, 1957, 106, 1-176.
3. Beck, S. J. *The Six Schizophrenias.* New York: American Orthopsychiatric Association, 1954.
4. Bergmann, G. The logic of psychological concepts. *Philo. Sci.*, 1951, 18, 93-110.
5. Bergmann, G. Theoretical psychology. *Ann. Rev. Psychol.*, 1953, 4, 435-458.
6. Bergmann, G. *Philosophy of Science.* Madison: University of Wisconsin Press, 1957.
7. Brown, M. H. and Bryan, G. Elizabeth. The altitude quotient as a measurement of intellectual potential. *J. clin. Psychol.*, 1957, 13, 137-140.
8. Burt, C. *The Factors of the Mind: An Introduction to Factor-analysis in Psychology.* New York: MacMillan, 1941.
9. Cattell, R. B. *Description and Measurement of Personality.* New York: World Book Company, 1946.
10. Cattell, R. B. Validation and intensification of the Sixteen Personality Factor questionnaire. *J. clin. Psychol.*, 1956, 12, 205-214.
11. Clark, R. A. and McClelland, D. C. A factor analytic integration of imaginative and performance measures of the need for achievement. *J. gen. Psychol.*, 1956, 55, 73-83.
12. Conger, J. J., Sawrey, W. L. and Krause, L. F. A reanalysis of Beck's "Six Schizophrenias." *J. consult. Psychol.*, 1956, 20, 83-87.
13. Conger, J. J., Sawrey, W. L. and Krause, L. F. Reply to Stephenson. *J. consult. Psychol.*, 1956, 20, 90-92.
14. Cronbach, L. J. and Gleser, Goldine. Assessing similarity between profiles. *Psychol. Bull.*, 1953, 50, 456-473.
15. Cronbach, L. J. and Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
16. Davis, A. and Hess, R. D. What about IQs? *J. Nat. Educ. Assoc.*, 1949, 38, 604-605.

17. DAVIS, P. C. A factor analysis of the Wechsler-Bellevue scale. *Educ. Psychol. Meas.*, 1956, 16, 127-146.
18. EYSENCK, H. J. Types of Personality—a factorial study of 700 neurotics. *J. ment. Sci.*, 1944, 90, 851-861.
19. EYSENCK, H. J. *Dimensions of Personality.* London: Routledge & Kegan Paul, Ltd., 1947.
20. EYSENCK, H. J. *The Scientific Study of Personality.* New York: MacMillan, 1952.
21. EYSENCK, S. B. G. Neurosis and psychosis: An experimental analysis. *J. ment. Sci.*, 1956, 102, 517-529.
22. FISKE, D. W. Consistency of the factorial structure of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, 44, 329-344.
23. FLEISHMAN, E. A. and HEMPEL, W. E. Factorial analysis of complex psychomotor performance and related skills. *J. appl. Psychol.*, 1956, 40, 96-104.
24. FRANK, G. H. The Wechsler-Bellevue and psychiatric diagnosis: A factor analytic approach. *J. consult. Psychol.*, 1956, 20, 67-69.
25. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, 52, 143-150.
26. GOLDSTEIN, K. and SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239.)
27. GUTTMAN, L. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, 19, 149-161.
28. HALL, C. S. and LINDSEY, G. *Theories of Personality.* New York: Wiley, 1957.
29. HORST, P. The uniqueness of configural test item scores. *J. clin. Psychol.*, 1957, 13, 107-114.
30. HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, 24, 417-441.
31. JASTAK, J. A plan for the objective measurement of character. *J. clin. Psychol.*, 1948, 4, 171-178.
32. KETTNER, N. W., GUILFORD, J. P. and CHRISTENSEN, P. R. A factor-analytic investigation of the factor called general reasoning. *Educ. psychol. Meas.*, 1956, 16, 438-453.
33. LAWLEY, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proceedings, Royal Soc. Edinburgh*, 1940, 60, 64-82.
34. LEVY, L. H. and DUGAN, R. D. A factorial study of personal constructs. *J. consult. Psychol.*, 1956, 20, 53-57.
35. LEWIS, D., MCALLISTER, DOROTHY E. and BECHTOLDT, H. P. Correlational analysis of the learning and relearning of four different tasks on the modified Mashburn apparatus. *J. Psychol.*, 1953, 36, 83-109.
36. LEWIS, D., MCALLISTER, DOROTHY E. and BECHTOLDT, H. P. Correlational study of performance during successive phases of practice on the standard and reversed tasks of the SAM complex coordinator. *J. Psychol.*, 1953, 36, 111-126.
37. LYKKEN, D. T. A method of actuarial pattern analysis. *Psychol. Bull.*, 1956, 53, 102-107.
38. MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, 14, 165-171.
39. MITCHELL, J. V. A comparison of the factorial structure of cognitive functions for a high and low status group. *J. educ. Psychol.*, 1956, 47, 397-414.
40. MOORE, T. V. The essential psychoses and their fundamental syndromes. *Stud. Psychol. Psychiat.*, 1933, 3, 1-128.
41. O'CONNOR, J. P., STEFIC, E. C. and GRESOCK, C. J. Some patterns of depression. *J. clin. Psychol.*, 1957, 13, 122-125.
42. RAO, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, 20, 93-111.
43. RAPAPORT, D., GILL, M., and SCHAFER, R. *Diagnostic psychological testing.* Chicago: Yearbook Publishers, 1946.
44. ROSENTHAL, IRENE. A factor analysis of anxiety variables. *Dissert. Abstr.*, 1956, 16, 376-377. (Examined in abstract only.)
45. SAKODA, J. M. Osgood and Suci's measure of pattern similarity and Q-technique factor analysis. *Psychometrika*, 1954, 19, 253-256.
46. SPENCE, K. W. *Behavior Theory and Conditioning.* New Haven: Yale University Press, 1956.
47. STEPHENSON, W. A reply to the reanalysis of Beck's "Six Schizophrenias." *J. consult. Psychol.*, 1956, 20, 88-89.
48. STEPHENSON, W. *The Study of Behavior: Q-technique and its Methodology.* Chicago: University of Chicago Press, 1953.
49. Technical recommendations for psychological tests and diagnostic devices. *Psychol. Bull.*, 1954, 51, 201-238.
50. THORNDIKE, E. L. The organization of a person. *J. abnorm. soc. Psychol.*, 1950, 45, 137-145.
51. THURSTONE, L. L. *Multiple-factor Analysis: A Development and Expansion of the Vectors of the Mind.* Chicago: University of Chicago Press, 1947.
52. TROUTON, D. S. and MAXWELL, A. E. The relation between neurosis and psychosis: An analysis of symptoms and past history of 819 psychotics and neurotics. *J. ment. Sci.*, 1956, 102, 1-22.
53. VANDENBERG, S. G. A comparative study of the stability of a factor structure. *Dissert. Abstr.*, 1955, 15, 1649. (Examined in abstract only.)
54. WELSH, G. S. and DAHLSTROM, W. G. *Basic Readings on the MMPI in Psychology and Medicine.* Minneapolis: University of Minnesota Press, 1956.
55. WITTENBORN, J. R. and HOLZINGER, J. D. The Wechsler-Bellevue and descriptive diagnosis. *J. consult. Psychol.*, 1951, 15, 325-329.

THE CONCEPTUAL AND TEST DISTINCTION OF NEUROTICISM AND ANXIETY

RAYMOND B. CATTELL

Laboratory of Personality Assessment and Group Behavior
University of Illinois

1. The Problem of Defining the Criterion

The strategically-planned researches examined in this article lead to two test batteries—one for degree of neuroticism and the other for intensity of anxiety. However, the details of their administration, standardization and effective use will be relegated to a later publication [17], though initial research forms are already available [10, 12]. Here we are concerned with the theoretical developments which made the tests possible and which are psychologically more important than the tests themselves.

Even when test measurement procedures are of primary interest, one can get nowhere without an adequate theoretical treatment of validity. And in the case of anxiety and neuroticism, the discussion of validity requires some fundamental thinking in the logic of science. Obviously, it is necessary to reject at once such practices as making an "anxiety scale" by heaping all kinds of anxious and pathological responses into a conglomerate questionnaire, or the notion that one can point to some single, directly-measurable, clinical criterion against which a test can be validated, as the industrial psychologist validates against an occupational performance. Some may assume that a psychiatrist's rating constitutes such a criterion. If general observation alone did not explode this last illusion, the recent work of Scheier [17], obtaining an r of 0 25 between the anxiety ratings on 80 patients by two psychiatrists selected for diagnostic reputation, would suffice to do so.

It is equally obvious that one cannot validate anxiety measures by showing that neurotics score more highly on them. We do not have two terms for nothing, and though neurotics *may* tend to have significantly higher anxiety, the better the correlation of the test with neuroticism, beyond a certain point, the poorer and more contaminated it is as a measure of anxiety itself. And while some have thus confounded anxiety with neuroticism, others have allowed the definition to slide into stress, and similar concepts, or, in a final masterpiece of confusion, into ergic drive strength itself. It is our task to find a sure methodological basis for separating anxiety, neuroticism, and, peripherally, stress and fatigue.

The multivariate experimental method, as distinct from the univariate manipulative experiment, does not start with guesses (postulates) regarding the unitariness of concepts, but shows by analysis of the covarying of phenomena what functional unities exist. Of course it still remains a semantic and experimental problem to decide which of the factor unities, A, B, C, etc., thus discovered are entitled to be interpreted and named anxiety (or anxieties), neuroticism, etc. It may be desirable for a time to represent the new operational entities only by symbols, or novel terms, thus avoiding the load of tangled misconceptions imported in such popular terms as "intelligence" or "anxiety".

Sooner or later, however, it will probably prove desirable to transfer these common terms, by a specific act of definition, to the concepts crystalized in the discovered factors, when further verified by experiment, as having the right to inherit whatever core of reality exists in the common term. If this transfer is to take place, however, the factor must be one that is central in a very comprehensive sample of variables commonly semantically attached to the idea in question—in this case anxiety. In the series of experiments summarized here, great care has in fact been taken to include representatives of all kinds of labelled anxious behavior according to professional usage, and, as it happens, a single major factor can be established in this area. The area is perhaps most succinctly, but comprehensively, covered by Thorne's clinical evaluation in the following condensed quotation. "Fear and anxiety are

associated with widespread psychosomatic syndromes characterized by feelings of *tension* and *irritability*, "band like" frontal and occipital headaches, dryness of the throat . . . gastro-intestinal symptoms, generalized fatigue and weakness, and "floating" or fixed (phobias) apprehension".[29, p. 258] Conceptual validation[13] of test measurements thus becomes no longer a matter of correlating any test with one special criterion, *e.g.*, an occupational success, but of correlating it with the *discovered designated major factor in the agreed general area of manifestation.*

However, medicine also knows an alternative approach to defining syndrome entities (correlation clusters or factors), namely by clinical recognition of pathognomonic symptoms, rooted in a certain etiology. Psychiatrists trained in this medical approach are inclined to depend on it a good deal and to argue that the search for unitary syndromes without knowledge of or regard for their causal factors is likely to be blind and confusing. Unless one knows, for example, that the patient has had a venereal infection, the initial onset of general paralysis would be quite difficult to distinguish from several other psychopathological syndromes.

A rather profound methodological issue is involved here, for fuller discussion of which the reader must be referred elsewhere[13] in which it is maintained that cross sectional recognition of form has to be prior to longitudinal perception of etiological change—as the single, well-focussed, cross sectional "frames" in a movie film must exist before movement of objects can be seen. In general, a pattern or entity *must be definable before its appearance and disappearance can be studied.* However, the methodological issues here include some aspects of the hen-and-egg conundrum, and it must be admitted that a methodologically-independent longitudinal, etiological basis for identification of functional unities exists. However, the latter requires, similarly, an application of factor analytic methods, in P-technique, in which situational causes and organismic effects are correlated longitudinally[4].

Except when the basic methodological considerations just mentioned permit longitudinal analysis, *and* with the developed statistical treatment of etiology just described, the medically trained psychologist is in danger of depending excessively upon identification by etiology. He is used to diseases that have very specific—and known—causes, whereas the psychological response patterns are more complexly combined and less specific to one causal situation. A germ of measles produces a measles syndrome; but an anxiety response pattern can arise from countless internal and external sources. Parenthetically, it is familiarity with the specificity of the majority of physical disease entities which has made it difficult for medicine to recognize the cause of rheumatic fever, or to accept Selye's concept of "nonspecific stress" disease.

It is methodologically preferable, even if not an absolute necessity, to use in psychology a more developed method for isolating response patterns, and to isolate these *before considering etiology.* Emphasis must be placed on recognizing and defining anxiety, neurosis and stress, by the refined and sensitive factor technique, directed to recognizing distinct unitary response patterns. The complete definition of the pattern will, of course, require supplementation by sequential, P-technique studies; incremental R-technique studies, as introduced below; and general etiological studies, centered on the recognized response pattern.

2. The Evidence for a Functionally Unitary Anxiety State and Trait

To the question whether there is *one* or *several* major factors in the area of responses considered to be anxiety responses, the unequivocal answer of at least seven researches on adequate population samples[2, 6, 8, 13, 15, 16, 17] is a *single* factor in behavioral tests (T-data), but *several* factors in questionnaire and life record (rating) variables, *i.e.*, in Q- and L-data[1][9, 13]. This paradox of the one and the many can be

[1]For economy of presentation we shall adhere to the practice started elsewhere[3, 6, 13] of designating the three media of observation: L-data (Life record, behavior *in situ;* by rating or time sampling); Q-data (Questionnaire, consulting room, self-report evidence); and T-data (Objective, behavioral, miniature-situational response in standard, non-questionnaire, test presentations).

resolved neatly enough, but only with sufficient attention to technical issues, as given below.

Let us first look at the single factor in objective behavioral tests, which thus strikingly confirms the clinician's unitary concept. It has been symbolized as U. I. (T) 24 in the proposed universal index series for personality and ability factors[5]. In Table 1, the loadings for this factor as found invariant (save for a due range of random error) in the two adult and two child populations collated earlier[6], and the Rosenthal[28] and Scheier[17] studies since, are set out. The table follows the usual conventions in defining patterns (see 6, 7, 8, and 13), namely, salients *common* to studies first, followed by salients found in single studies only; reversal of verbal labels for tests where signs are negative; and identification of experimental test by a master index number through which the actual test can be obtained[10] for further experiment.

TABLE 1. LOADING PATTERN OF U. I. (T) 24: GENERAL ANXIETY

Master Index Number	Manifestation (Variable) Title, with Present Direction	Factor: Series Identification, Sample Size and Research						
		1	9(–)	12	5	1	1	1
		500	250	184	100	70	80	80
		(8)	(8)	(16)	(2)	(28)	(17)	(17)
	Common to Several Studies							
211	High Annoyability	52	39	21		49	10	
219	Willingness to confess frailties	–42	–19	–43		–38	–36	–32
152	Tendency to agree	24	63	49		26	18	
21	Absence of questionable preferences	33		09	36			
116	High ratio hard-headed to sentimental-indulgent attitudes	28		48		33		
205	High emotionality of comment	35	13	17				
108	Low confidence in untried performance	–11	–03	–20			–25	–26
276	Good grades (intelligence out)			18	40			18
274	High verbal ability			19	55			
29c	Critical severity						12	24
	Found in one Research							
	Low hand steadiness (tremor)					22		
	Lesser saliva volume							–38
330	Good ethical choices				44			
327	Logical inconsistency of responses	51						
	Higher ratio acquaintances to friends						30	
294	Low "contact with reality"	–43						
	Slower tempo						29	
	Greater sex erg tension by objective attitude tests						42	
	Less increase of heart rate after startle						–25	
	Questionnaire and Rating Variables							
	16 P. F. Q_4 Ergic Tension	74				70	62	47
	16 P. F. O Anxious-depression	66				84	33	33
	16 P. F. Q_3(–) Low Will (Self Sentiment)	–60						60
	16 P. F. C(–) Low Ego Strength	–56				–46	–47	–20
	16 P. F. L Protension (Paranoid)	54						25
	16 P. F. M Alutia	46				24	18	18
	16 P. F. H(–) Threctia	–33					–36	–18
	Checklist of Anxiety Symptoms					85	56	57
	Anxiety rating by psychiatrist*						39	

*Mean of 2 psychiatrists. Only this factor was significantly loaded in the ratings of both psychiatrists.

The identification of this sufficiently replicated pattern with the concept of general anxiety is made, first, by virtue of its containing the tests of such diverse anxiety manifestations as irritability, emotionality and tension, lack of confidence, self criticism, tremor, reduced salivary secretion, etc. The significant loading in "*less* increase of heart rate by startle" is presumably due to the anxious person having already, like Macbeth, "supped full of horrors". Secondly, it is made by reason

of the very large loadings it contains in those questionnaire factors that have anxiety content; in the anxiety *symptom* check list; and in the psychiatrist's rating. In both cases the logical proof requires *also that these do not appear* together in any other of the twelve to eighteen personality factors[13], and this condition, we find, is also met.

A number of important and decisive inferences can be made from the special combinations of behavior found in this pattern, notably the combination of tension loadings with good school grades and good ethical choices; but space compels us to transfer the discussion elsewhere[13] and to attain here simply the demonstration of a consistently replicated, complex pattern. However, before leaving this pattern, to deal with its further relations— to Q- and L- data patterns, and to neuroticism— the theoretically very important fact should be mentioned that it has been established not only as a pattern in individual differences (R-technique) but also as a pattern of coordinated change *within* the individual (by incremental R-technique and P-technique). In other words it is a dimension defining a *state* dimension as well as a *trait*. Indeed the individual differences in trait *may* be nothing more than relatively permanent differences in state[7] level.

The pattern found for the incremental R-technique (factoring of *differences* between two anxiety stimulus levels) is set out elsewhere[17] and the match is very good. It adds: higher systolic blood pressure, distractibility, and fluency in self criticism. Parenthetically the reader should note that when we say factors in independent studies "match" we do not mean merely by impression, and inspection of a table, but that: (1) the two independent simple structures, obtained without rotation prejudiced by any theories, has been checked by Bargmann's test for significance[1], and by the secondary criterion that at least 60% of variables must fall in the hyperplanes, and, (2) the pattern agreement of two researches has been checked on actual common marker loadings by the Cattell-Baggaley "*s*" index for significance of factor similarity[14].

In the completeness of this demonstration of the functional unity of the anxiety pattern, however, one lacuna still exists. The riddle remains that we have not yet been able indubitably to recognize the factor in P-technique studies[7]. This is largely due to an accidental deficiency of the necessary markers for identification in the studies concerned. The resultant uncertainty consists in that the U. I. (T) 24 match could be either with U. I. P. F. 9 (–), previously labelled General Autonomic Activity, or with U. I. P. F. 7 (–) the Basophil-neutrophil ratio pattern[7]. The former is more probable, and makes better sense theoretically in that it contains the chief physiological variables of autonomic over-activity; but an important study remains to be carried out here to give that dependable identification of the P-technique pattern which is now needed for theoretical advance.

3. The Evidence for the Distinctness of the Neuroticism Dimension

Now the especial value of the above location of the "general anxiety factor" in objective (and physiological) measures— as distinct from the identification in the latter patterns (below) occurring in Q-data only— is that it permits a clear distinction of anxiety and neuroticism. For a factor has been found in the same T-data series, simultaneously, in several studies, indexed as U. I. (T) 23, which there is good reason to identify with neuroticism.

First, as to the sufficiency of confirmation of the U. I. (T) 23 pattern in itself: as early as 1955 it had been verified[6] in seven independent studies— five in our laboratory and two in Eysenck's— with groups of considerable diversity: adults, children, psychotics and neurotics (See Table 2). Blind simple structure rotations have produced it three more times since 1955[15, 17, 28] and one of these researches was an incremental R-technique, showing that over a two month interval people rise or fall, slightly, simultaneously, on the key variables on the pattern (Column 1, Table 3). Conceivably we are again dealing with trait (individual) differences that are largely state (occasion) differences, but, compared with anxiety, the function fluctuation is apparently less and over longer periods.

The proof that this dimension is distinct from anxiety is twofold: (1) The content, *i.e.*, saliently loaded variables, is quite different, and (2) in practically every research in which U. I. (T) 24 has appeared, U. I. (T) 23 has also been found at the same time. Therefore, U. I. (T) 23 could not be U. I. (T) 24 reappearing in a different pattern of unrelated variables in a different research. Indeed, in the simple structure plots of 23 and 24, it is possible to see their hyperplanes very distinctly and to determine their relation (angle of hyperplanes) with considerable exactness. (They prove to be slightly but significantly positively correlated.)

Compared with the demonstration of *unitariness*, and *distinctness from anxiety*, the proof that U. I. (T) 23 also *corresponds to the clinical conception* of neuroticism encounters some complications. The initial proof that the variables in U. I. (T) 23 distinguish neurotics from normals, better than any other known tests, we owe to Eysenck, Himmelweit, Furneaux and others of the Maudsley group[20, 21]. But the present writer cannot accept that Eysenck's principle of criterion rotation is a self-sufficient principle, and, indeed, it is subject to the same criticism as the obscure practice of validating anxiety scales in terms of differences of neurotics and normals, mentioned above. In criterion rotation, it is true, the crass assumption is not made that the criterion is simply "anything distinguishing neurotics from normals", for the extra condition is required that one be able to produce a *factor* pattern corresponding exactly (except for chance error) to the *difference* pattern. Nevertheless, with present techniques, if neurotics differed from normals *both* in neuroticism *and* in anxiety level, a complex factor confounding the two might be mistaken for "neuroticism".

As it happens, the certainty (statistical significance) with which simple structure has been obtained in the five researches in our laboratory, and the simultaneous independent demonstration we have made of the anxiety factor, show that Eysenck's criterion rotations did *not* happen to confuse two factors. U. I. (T) 23 is not U. I. (T) 24, and it remains the factor most powerfully distinguishing neurotics from normals. But is it exactly what the clinician calls neuroticism?

Since the largest study[20] was on severe and hospitalized neurotics, the contaminating dimension next most open to suspicion would be psychoticism. Eysenck, however, has always been very alert to this possibility, and has compared neurotic and psychotic groups, with results favorable to the theory that neuroticism and psychotism are entirely distinct directions of maladjustment[21]. Our own results indicate a factor U. I. (T) 25 apparently identical with the pattern he obtains distinguishing psychotics from normals. It is not, as yet, replicated exactly enough, on a sufficient number of variables, in a sufficient number of studies, to stand alongside U. I. (T) 23 and 24 in quality of definition, and reliability of pattern measurement. But it is clear enough to show that we are encountering a dimension of behavior distinct from neuroticism and anxiety, and peculiarly distinctive of psychotic groups when contrasted with neurotics and normals.

Thus if the separation of groups designated neurotic, psychotic and normal is the pay-off, U. I. (T) 23 *is* neuroticism, and for, the rest, cynics may ask "What and where is this agreed concept, among clinicians, of neuroticism, which U. I. (T) 23 *could*, perhaps *more exactly*, depict?" The consensus on neuroticism definitions is indeed a very limited one; but perhaps it covers "excessive conflict of ego, id and super ego, without solution by either successful repression or stable expression." The doubt which caused us[5, 6, 8, 10, 13] to call U. I. (T) 23 "Low Neural Reserves *vs.* Ego Strength", instead of "Neuroticism", arose from the fact that signs of conflict are not particularly conspicuous in the variables, but, instead, there is chronic fatigue, lowered coordination, and inability to cope with behavioral problems—which might be the *consequence* of prolonged conflict, rather than the conflict itself, or even the result of physical illness. As Thorne points out, a pattern characterized by exhaustion, mixed symptoms, overwroughtness and inability to concentrate on immediate stimuli has been clinically recognized, *e.g.*, in Spencer Cowles' "acute

neural exhaustion state", representing an organismic failure to achieve integration under stress as in the classical "neurasthenia".[2]

It is this generalized breakdown which is almost undoubtedly the clinical expression of extreme U. I. (T) 23 measures. The particular form of the neurosis— the specific pathognomonic signs and relationships— will depend on the specific dynamic conflicts, and on those temperamental and general personality factors, *e.g.*, surgency-desurgency (F), cyclothymia-schizothymia (A), tendency to anxious depression (O), which express the essential personality [13] as it persists through the neurotic or psychotic translation along factors U. I. (T) 23 and 25. It has been with the intention of retaining the meaning for U. I. (T) 23 (–) as a more generic "exhaustion *vs. immediate* capacity to mobilize" that we have generally avoided the designation 'neurosis' and used [6, 13] "lowered neural, hypothalamic reserves (incapacity to mobilize) *vs.* ego strength".

TABLE 2. THE LOADING PATTERN OF U. I. (T) 23 (–), NEUROTICISM *vs.* EGO STRENGTH

Master Index Number	Manifestation (Variable) Objective tests Title, With Direction in Present Factors	Factor: Series Identification, Sample Size and Research 10 370a (See (6))	8 370a (6)	7 100a (6)	7 100a (6)	5(–) 184c (16)	1 105a (20)	1 98a (21)	4 80 (17)	4 80 (17)
	Common to Several Studies									
2	High motor rigidity	15	29	58	20	37	42		37	25
41	Poor two hand coordination	–02	–20	(15)	–37	–36		–42		
42	High body sway (and suggestibility)	20	13	26	12	10	64	12	15	33
120	High ratio of inaccuracy to speed			20	20	36	40		–38	–37
28, 223	Low "dynamic momentum" (endurance of difficulty)	01		*12*	*36*		*50*	*16*		
151	Slower decision on particulars than principles			–31	–27				–04	
32, 1	Larger P. G. R. to mental than physical threats		07	31						
101, 148	Much excess of aspiration over achievement			29	02	56	17			
162	High ratio of shock to normal performance			–26	–46	–26				
1	Low fluency	00	–09			–10	–40			
167	Poor immediate memory			–14	(01)	–01	00	–33	–13	
216	High suggestibility about shock					17		15		
51	Low carefulness of observation		–67							–41
	Found Essentially in One Research									
	Much mind wandering (Inability to concentrate)								41	
	Few figures added correctly								–39	–23
	More distraction (by esthetic temptation)								59	
	Less spread than localized effort								–58	
	Less renunciation								–57	
	Less inhibition from failure								–41	
	Low metabolic rate change with stimulation		–57						–57	
	Low salivary output								–34	
	High P. G. R. absolute resistance								31	
	More pupil dilation in stress									38
	Avoidance of stronger smells									35
	Ratings, Questionnaires									
	16 P. F. Low Ego Strength C(–)			41					21	
	16 P. F. Anxious Depression O								23	
	16 P. F. Lack of Will Control Q_3 (–)									26

Parenthetically, the reader familiar with the standard factor representation will recognize that U. I. (T) 23 is presented here in the *opposite* to the usual *positive* direction, since, U. I. (T) 23 (–) is the "neuroticism or ego weakness" direction, with

[2] Personal communication.

which we are here concerned. (Experimenters should thus be alert to the changed direction of loading and scoring from the standard U. I. (T) 23 presentation.) The central feature in the two dozen varied performances above is surely a form of fatigue (rigidity has been shown to increase with fatigue), a distractibility and a poor capacity to mobilize appropriate immediate responses (which, by the factoring out, is proved to be distinct from low intelligence). Premature crystallization of an unduly specific hypothesis about this incapacity is perhaps best avoided by the label "low neural reserves" or "debility", defining an effect on all kinds of central nervous system performance: cortical, cerebellar and hypothalamic. Evidence from P-technique and incremental R-technique will show that though this can be labelled a *kind* of fatique, it is of longer duration and different in nature from ordinary diurnal fatigue, which is demonstrably a statistically distinct factor, though lateness of day also contributes, slightly, positively, to this "over-wroughtness" factor (see below).

To complete this evidence of functional unity of U. I. (T) 23, the factor has been demonstrated by Scheier[17] in analysis of increment scores on the same markers (with students, respectively well rested and after hard work, immediately before examinations) and independently by Cattell[7] and Haverland[24] in P-technique as a factor initially labelled "Overwroughtness *vs.* Untapped Reserves" (P. U. I. 8). The alignment of these patterns is shown in Table 3. It is not yet as good as could be

TABLE 3. GENERAL NEUROTICISM (U. I. (T) 23 (–)) *vs.* EGO STRENGTH AS A STATE FACTOR

Manifestation Title, With Direction in Present Factor	Factor: Series Identification, Sample Size and Research		
	7 Incremental 80 cases (17)	1 P-tech. 1 case (See (7))	10(–) P-tech. 1 case (24)
Higher disposition rigidity	32	54	17
Slower reaction time, especially unwarned	11	44	14
Smaller myokinesis	–24	–44	
Lower fluency		–07	–38
Lower endurance			–29
Reduced hand steadiness	–11		–42
Short hours of sleep previous night		51	26
Jokes rated unduly funny		55	
Small movements in myokinesis		–44	
Lower ratio accuracy to speed	–38		
Higher sway suggestibility	10		
Lower ego strength; 16 P. F. Factor C(–)	–31		
Greater id tension; 16 P. F. Factor Q_4	16		
Hectic activity (obs. rating)		29	
More acceleration by competition	24		
More response error in complex reaction times	20		

desired, the pattern similarity significance centering on P = .03 for the three matches. However, as stated, it is at any rate important that the pattern is clearly distinguishable from the diurnal fatigue factor in the same series. This study shows that the pattern elements found in the factoring of *individual differences*— high rigidity, sway suggestibility, low fluency, low endurance, low effectiveness of concentration in calculation, etc., has elements which also vary together in one person with time and circumstance. Further, it adds that the "neurotic" pole is also connected, nevertheless, with social overactivity and excitability.

The essential nature of U. I. (T) 23 (–) might, in summary, be designated as an inability to coordinate and control responses of almost all kinds in *immediate* activities. One is strongly tempted to call it "lack of integration", but this would invite confusion, for there is no evidence of lack of true *dynamic integration*, *i.e.*, of harmony of goals and self-sentiment values, but only of inability to effect *immediate integration* in muscular control, psychological impulse, memory, the overcoming of fatigue, etc. Indeed, all the evidence points to this other kind of lack, *i.e.*, of dynamic

integration, being synonymous rather with U. I. (T) 24, Anxiety, as seen below. The word *immediate* is therefore italicized to emphasize that the "lack of integration" is not of a wider character, visible in purposes and values, but simply represents inability to control immediate motor activity, thoughts, productivity and impulses, in mobilization toward immediately accepted goal. Now this may well be the correct meaning of "ego strength", in Freud's description [22], though some clinicians have allowed the meaning to drift into "general dynamic integration".

In favor of this identification of U. I. (T) 23 (–) with "debility *vs.* ego strength" is the fact that its only consistent correlation with a questionnaire factor has been with 16 P. F. factor C (–), independently denoted, earlier, as Ego Strength. Doubt about complete identity, however, arises from the facts: (1) that the correlation has not been as high as would be expected, with attenuation correction, for a complete identity, (2) that this correlation *might* really be no higher than would be expected as part of the correlation of U. I. (T) 23 (–) with U. I. (T) 24, which shows itself, secondarily, as a series of quite low correlations of U. I. (T) 23 (–) with the questionnaire factors, Q_4, Q_3 (–), O and C (–) which high$_3$y $_4$oad U. I. (T) 24 (See Table 1), and (3) that questionnaire factor I (Premsia or Sensitivity) correlates as highly with U. I. (T) 23 (–) as does C (–), in certain samples. Granting this much uncertainty, the favored hypothesis is sti that C and U. I. (T) 23 are identica , and a ike appropriately interpreted as Ego Strength *vs.* Debility (or Generalized Neuriticism or Impaired Immediate Ability to Mobilize.

As stated above, this neurotic debility is not to be considered as the specifically recognized neurotic behavior *per se*, which is determined partly by the specific precipitating environmental stress and the pre-neurotic personality. It consists instead of the essential ego weakness, visible in the multiplication of defense mechanisms, in instability, in a sense of inadequacy, and in a preponderance of dissatisfied emotionality and impulsiveness, as brought out most clearly in the L-data (life record) factor [3, 13] pattern for C (–). The first alternative to explore, if the hypothesis of *identity* of C and U. I. (T) 23 fails, is that U. I. (T) 23 (–) is merely a more restricted "exhaustion" factor arising from C (–); but this is less likely.

4. Anxiety Origins and Expressions Shown in the Second Order Factor

Returning now to the anxiety factor, we have to investigate the paradox that though there is one U. I. (T) 24, there are *several* personality factors, with discernible anxiety content, in L- and Q-data. Fortunately, we are granted a rapid denouement. The recent, careful, second-order factor analysis over the primary factors in personality shows that both in L-data [13] and Q-data [9], these particular primaries, covering diverse aspects of such behavior, come neatly together in a second-order factor, which has been called Anxiety *vs.* Dynamic Integration. In diminishing order of loading (combining Table 4 and L-data results in [13]) the factors are: Q_3 (–) Lack of will control or defective self sentiment formation; Q_4, Ergic tension (Id pressure); O, Anxious depression and guilt; C(–), Ego Weakness; L, Protension or paranoid trend; and H(–) Threctia (or susceptibility of sympathetic system to excessive responsiveness to threat) together with one or possibly two less constant factors shown in L-data in [13] but not in the Q-data of Table 4. Certain conglomerate "anxiety scales" and symptom check lists succeed [28] in falling largely in this second-order

Table 4. The Second Order Anxiety Factor Appearing Among Primaries in Q-Data

Symbol	Factor I Anxiety *vs.* Dynamic Integration	Loading
Q_3(–)	Lack of Will Control* or Poor Self Sentiment Integration	.66
O(+)	Anxious depression or Super Ego Susceptibility	.62
Q_4(+)	Ergic Tension or Id Pressure	.62
L(+)	Paranoid Trend (Protension)	.54
C(–)	Lack of Ego Strength	.53

*The titles used are those of the 16 P. F. Handbook [18]. To make easy, consistent reading all loadings have been made positive, by reflection, in these factor lists. The reflection is indicated by the sign on the left and the change of title to the opposite pole, e.g., "Lack of . . ."

factor, and are evidently varying mixtures of the primary personality factors just defined. Inspection of the original rating[13] and questionnaire[18] items will show that these primaries cover all primaries that could conceivably be interpreted as anxiety, except possibly primaries I, (Sensitivity, or Premsia) and F(–) (Desurgency), whose exception is discussed elsewhere[13].

Although the second-order structure, analysed on three independent populations, is very convincing[13], the evidence for such a second-order factor in Q– and L– data, *and* for its identity with the first order objective test anxiety factor U. I. (T) 24, does not rest on factor analysis alone. For, additionally it has been found that: (1) These same six factors all show very similar change scores with age, being high in adolescence and dropping to early maturity (differing in this from other primary factors)[18]. (2) In all studies[8, 17, 28] where objective tests and questionnaire factors have been measured together, the above *six* correlate directly and very substantially with U. I. (T) 24 and *only* with U. I. (T) 24— except for C, as stated above. They do so with the same sign and magnitude as would be expected from the identification of the second order factor with U. I. (T) 24 (See Table 1). This includes both absolute and incremental factorings. (3) Important sub-tests of U. I. (T) 24 and questionnaire responses have been shown to change *simultaneously*, as their unity requires, in students tested at and remote from examinations, and in air pilots before and after threatening flight experiences. (4) These same six factors have an identical pattern of correlation with other major entitites (Tryon's criterion of functional unity[30]), notably with measures of interest in career, in self sentiment and in need for security[13].

Granted this demonstration of a functional unity among certain primary factors, assessed either from L- and Q-data, at the seecond order level, and its identification with U. I. (T) 24 in T-data, how do we explain (a) the non-appearance of six corresponding primary patterns directly in T-data, and (b) the nature of the relation of the six Q- and L-primaries to the unitary second-order or (T-data first order), which is, after all, the *essential* anxiety unity running across Q-, L- and T-data?

One explanation why the primaries do not show up in T-data (and that therein we pass instead directly to the second-order unity) could be that the T-data observations are concerned with very general behavioral and with physiological responses, not embedded in the more specialized and particularized ramifications of life purposes, attitudes and self concept values capable of yielding the more refined structure found in L- and Q- data. The change from a single factor to a six factor group would thus be explained analogously to the transition from a lower to a higher power on a microscope, resolving a single blob of bacteria into several members, but this description still leaves us in need of a hypothesis to account for the primaries adhering together in this way. Confirmed, empirical, second-order factors are still too few to have yielded any generalization about the *usual* causal dependence in such cases; but we know from statistical logic that the elements loaded in a factor can be either causes or consequences of the factor itself. We also know that probability is on the side of their being consequences, since it is hard to imagine a lot of essentially independent causes being as highly correlated as would be necessary for them simultaneously to load one factor. However, the only sure answer is to be obtained through supplementing the factor analysis by controlled experiment. In the following hypotheses to explain the origins and consequences of anxiety, developed more specifically by present quantitative findings from hypotheses put forward seven years ago on general psychoanalytic and clinical evidence[3], some loadings of primaries in U. I. (T) 24 are made causes and some effects, according to ancillary indications.

1. The hypothesis[22] that anxiety will be proportional to id pressure seems to be born out by the Q_4 component, one of the largest of the six. However, our present knowledge of Q_4 is in its infancy, to the extent that it may have to be considered *total ergic tension*[13] rather than *repressed* ergic tension only. This still does not identify anxiety with "general drive"; but says that only *one of the components* in anxiety is total undischarged ergic tension. High total ergic tension, from whichever of the standard causes[13], could produce anxiety directly, by threat to the ego, *i.e.*, to

the master need to maintain unity, or because to have more needs in the environment is to have proportionately more danger of loss, frustration, suppression and repression. This hypothesis therefore states that anxiety is a derivative of ergic tension, and not that all ergic tension is transformed into anxiety or that all anxiety is ergic tension.

2. As hypothesized [2, 22] ego weakness is associated with anxiety, as shown by the consistent C(–) and U. I. (T) 23 (–) correlations. But this may either confirm the more specific hypothesis [3] that fear of overthrow of the ego generates anxiety, or the converse causation that high anxiety (especially in the developmental period) encourages ergic regression and repression, and, therefore, *impoverishment* of ego development. Or *both* causal directions could operate, as far as this present evidence of association goes.

3. The powerful role discovered for the guilt-depression factor, O, may or may not support the hypothesis, central, in different ways, in the anxiety theories of Freud [22] and of Mowrer [27] that super ego threat is responsible for much anxiety. There seems little doubt, from the content of O (emphasis on fear, guilt, inadequacy, dependency, and high standards) and from its appearance in extreme form, as a psychotic "guilt" factor [13], that it represents either a super ego development or, more likely, a dependence which makes the individual prone to marked super ego development. Until this is cleared up, and the relation to the "effective" super ego pattern G (in the 16 P. F. factors) clarified, we can speak only of a special "depressive anxiety" component (distinguishable from simple depression or desurgency, F(–), which is not involved in U. I. (T) 24).

4. No previous hypothesis accounts for the loading in factor L, protension or paranoid trend. L's description by content, and its effect in group dynamics [13], has been: inner tension, suspicion and a special tendency to the projective (in the true psychoanalytic sense) defenses. The false, dangerous, paranoid relation to society may *generate* anxiety, (through partial insight) or, more likely anxiety, as a social insecurity of unspecified origin, may cause either (a) paranoid, defensive over-self-evaluation or (b) cognitive deterioration through prolonged anxiety, in accordance with physiological theories of paranoia.

5. The excessive preoccupation of many clinical psychologists with elaborate theories *per se* perhaps accounts for neglect of the obvious quantitative hypothesis that if a capacity to be apprehensive of future dangers is part of anxiety then the individual's anxiety level will tend to be proportional to (a) the situational dangers and (b) the constitutional strength of his proneness to timidity! It may be, however, that timidity of disposition is not the same as level of fear erg tension. Indeed, when ergic tension strengths have been measured along with the present variables, as in the work of Rosenthal [28], Scheier [17] and Wenig [19], no significant positive relation has been found between fear of real dangers (ergic tension of fear in relation to standard dangers) and anxiety. The only ergic tension consistently found substantially correlated with anxiety has been that of the sex drive [13, 17, 28], fitting Freud's theory that sex is paramount among the undischarged drives (our Q_4 component); and possibly, also, his theory of "the psychosexual constitution."

In spite of this lack of relation of anxiety to the fear erg, found as a dynamic, 'ergic tension' factor [13], we find a very definite relation (Table 1) to timidity as a temperamental factor, H(–). Dimension H was initially, descriptively, called *Boldness vs. Shyness* (and is probably represented also by Guilford's S factor [23] according to an r of 0.85 recently obtained between H and S), but in view of increasing evidence of autonomic associations, specifically of parasympathetic blocking of sympathetic fear response, in H+, it has been renamed, by hypothesis, *Parmia vs. Threctia*. The threctic pole, H(–), defining "ready autonomic responsiveness to threat", correlates 0.33, 0.18 and 0.36 (in samples of 500, 80 and 80 respectively) with anxiety, indicating that a disposition to timidity (for H is highly inherited [13]) is not a negligible consideration in the genesis of anxiety. It is probable that an individual's scores on both H(–) and O represent a product of his proneness to situational anxiety and the strength of threat at the time, whereas C(–), L and Q_3 (–) would be expected to have little relation to the situational threat. The H (-) loading thus supports our "situational timidity" theory.

6. The most highly loaded primary in the second-order factor (Table 4 and [13]) is Q_3(–), and it rivals Q_4 for highest mean loading (correlation with) U.I. (T) 24 in Table 1. Corrected for attenuation these loadings *could* indicate that Q_3(–) is virtually the essence of the second order factor. Q_3 has been interpreted, through its content and its many criterion associations [18], as a will control factor representing the organized self sentiment in action. That is to say, it represents the strength with which the individual strives for behavior integrated according to a consciously approved (but not unrealistically idealistic) self concept.

That Q_3 should, factor analytically, be the cornerstone among the six primaries supporting the general anxiety factor, is of great psychological interest. It suggests that the effectiveness of the Q_3 striving for total integration is limited by the other conditions— degree of ego weakness, C (–); super ego pressure, O; total ergic tension demands, Q_4; etc.,— and that anxiety is nothing less than the degree of failure in the final dynamic integration. As dirt is matter out of place, so anxiety is ergic tension out of control. Or, to refine the simile, anxiety represents *the extent of the threat to the organism's effective dynamic expression as an integrated system*, in running equilibrium. If so, the true opposites in our conceptions should be *anxiety vs. good dynamic integration, not* neuroticism *vs.* integration. Neuroticism, U. I. (T) 23 (–), must therefore be conceived *not* as a lack of dynamic integration, but only as that special contribution to mal-integration which arises from ergic regression followed by defective build up of ego strength. The resulting chronic, systematic failures to face ergic and super ego demands, accompanied by constant anxiety of ego overthrow, constitute the specifically neurotic contribution to anxiety.

In so far as it is necessary to relate these findings to current psychiatric divisions of anxiety into "free" and "bound" anxiety, and into central (intrinsic or inherent) and peripheral (or situational) anxiety, it should be noted that in all except factors (H–) and Q_4 we are dealing with anxiety inherent in the system—a part of the make-up of the personality. The strengths of factors H(–) and Q_4, by contrast, represent in each case a function of interaction of the characteristics of the individual *and* the situation. H(–) represents the degrees of arousal of apprehension, as a product of the individual's timidity and the strength of environmental threat: Q_4 represents the *total* ergic excitation, as a product of the individual's excitability and the level of stimulation in the environment. By the nature of factor analysis these expressions in the individual of combined susceptibility and stimulus intensity appear in present techniques as single factors, though there are ways[4] of fissioning them into organismic and environmental factor components if these exist. The fundamental equation for an ergic tension factor score[13, p. 621], breaking it into situational and other components, applies to H(–) and Q_4 here as to any other ergic tension. A calculation from Table 4 [and 13] suggests that the two dispositional-situational sources, H(–) and Q_4, contribute little more than 1/3 to the total individual difference variance in anxiety; but this must be rough until the second order loadings are more exactly evaluated and the situational terms in the equation determined.

As to the psychiatric concepts of free and bound anxiety, there exist two factor dimensions to which they could be attached. Probably the obsession-compulsion neurosis, with which the idea has been principally associated, is not common enough, in those "normal" population samples in which factors have been established, to yield a factor of large variance, and probably K (–) and M (–), both of which have low loadings in the second order factor, best represent this obsessional ingredient. But, granting that binding is defined to mean also socially apt and useful habit discharge of "anxiety", then Q_3, with all its emphasis on socially approved behavior and self control, is the major measure of the individual's success in binding anxieties through effectively stabilized habits. Since Q_3, K and M are all *negatively* loaded in the major second order factor, this makes the anxiety factor (and U. I. (T) 24) in our discussions synonymous with *free* anxiety. Indeed, by the structure shown here, bound anxiety, *i.e.*, successful, stable, habit expression, is the very antithesis of anxiety, for it is the integrated energy discharged in a dynamic equilibrium through Q_3. This suggests that clinical conceptions would be clearer if "drive" were substituted for the expression "bound anxiety". For though drive may at some phase have expressed itself as anxiety, when bound it is certainly no longer anxiety, and no more entitled to be called anxiety than to be called any other drive, *e.g.*, sex, aggression, from which it may have been derived before passing into the bound state.

Summary

1. A general factor of anxiety— U. I. (T) 24— is found, at the first order level, on objective measures of anxious behavior of all kinds and as a second-order factor among questionnaire (Q-data) and life-record-in-situ (L-data) primary personality factors.

2. A factor of general neuroticism, or better, neurotic debility vs. ego strength, U. I. (T) 23 (–), has repeatedly been found to distinguish neurotics from normals, and to be distinct from the anxiety factor (though slightly, significantly correlated with it in most populations).

3. Both the anxiety factor and the neurotic debility factor have been demonstrated as patterns in *states*, as well as *traits*, *i.e.*, as unitary factors in incremental measures from day to day.

4. The L- and Q-data primary personality factors forming the second order anxiety factor, are Q_3 (–), Q_4, O, C (–), L, H (–) and possibly K (–). A theory explaining these as causes and consequences of anxiety has been put forward through the checking of these findings with earlier hypotheses.

As a result of the above, batteries for the measurement of anxiety and neuroticism have become available to the clinician. In objective, behavioral tests, essentially ten sub-tests are available for each battery[10]. All twenty *can* be, and three *have* to be, individually administered. Six to ten sub-tests constitute an absolute minimum for reliable measurement of each factor with present knowledge as to available high-loaded tests. Research needs urgently to be directed to evolving tests with higher correlations with U. I. (T) 23 and 24. It is hoped that the current work of Scheier on the anxiety factor will result in the discovery by 1958 of such a battery, exceeding present saturations, and still of no more than one hour's duration. According to the method of estimate proposed elsewhere[11] the existing 1955 batteries for neuroticism and anxiety have construct validities (correlations with the pure factors) of about +0.8.

Meanwhile, U. I. (T) 24 can also be measured through the questionnaire medium, and an Anxiety Scale[12] has therefore been constructed from the forty best loaded items, combining the Q-factors $Q_3(-)$, Q_4, O, C(-) and L with due regard to the weights required for the most accurate estimate of the second-order factor. This Anxiety Scale is further divided into a score from 'overt', manifest questions, plainly dealing with anxiety symptoms, and a score from 'cryptic' questions which load as well on the general anxiety factor as the others, but do not seem to the patient so obviously to deal with symptomatology. However, the Anxiety Scale, requiring less than ten minutes to administer, is intended primarily to yield a single anxiety score. If good reliabilities are required for part scores, as a diagnosis of anxiety *sources*, according to the above theories, it would be desirable, at least in research, to supplement it with the 16 P. F. scores of the patient on the full test length for the primaries $Q_3(-)$, Q_4, etc. This also would have the advantage, if the above hypotheses can be considered reasonably verified by the present data, that the particular scores on the factor measures, in the individual case, will throw light on the etiology of his anxiety state, on the chief dynamic conflict sources of his anxiety, and the manner in which he is channeling its discharge.

Certainly the provision of distinct, factor-defined and reproducible tests, for anxiety and general neuroticism (the former either in a single objective battery or in a brief questionnaire permitting breakdown into primaries) should make possible a wide array of meaningful experiments into which the clinician could not previously confidently enter. The existing evidence[6] suggests that these batteries[10] are already sensitive enough not only to give significant findings for inference from individual differences, but also, like a clinical thermometer, to register significant changes in one individual during the course of therapy. From a study of curves of anxiety production and recession, at various critical stages of therapy, far greater insight and control of the process than now exist might become common in the art of therapy. And, from the standpoint of research into general laws, these instruments should permit clear-cut answers in response to clear-cut hypotheses.

Bibliography

1. Bargmann, R. 'Signifikanzuntersuchungen der Einfachen Struktur in der Faktoren-Analyse.' Sonderdruck, Physica-Verlag, Wurzburg, 1954.
2. Brogden, H. E. A factor analysis of forty character tests. *Psychol. Monogr.*, 1940, 52, No. 3 (whole No. 234), 39-55.
3. Cattell, R. B. *Personality: A Systematic Theoretical and Factual Study.* New York: McGraw Hill, 1950.
4. Cattell, R. B. *Factor Analysis.* New York: Harper Bros., 1952.
5. Cattell, R. B. A universal index for psychological factors. Adv. Publ. No. 3, 1953, Laboratory of Personality Assessment, Univ. of Ill., Urbana.
6. Cattell, R. B. The principal replicated factors discovered in objective personality tests. *J. abn. soc. Psychol.*, 1955, 50, 291-314.
7. Cattell, R. B. The chief invariant psychological and psycho-physical functional unities found by P-technique. *J. clin. Psychol.*, 1955, 7, 319-343.
8. Cattell, R. B. Psychiatric Screening of Flying Personnel. Personality structure in objective tests—a study of 1,000 Air Force students in basic pilot training. USAF School of Aviation Medicine, Project No. 21-0202-0007, Report No. 9, Feb., 1955, 1-50.

9. CATTELL, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, 20, 411-418.
10. CATTELL, R. B. The Objective-Analytic Personality Factor Battery. IPAT, 1604 Coronado, Champaign, Ill., 1956.
11. CATTELL, R. B. Formulae and table for obtaining validities and reliabilities of extended factor scales. *Educ. Psychol. Meas.*, 1957 (In Press).
12. CATTELL, R. B. The I. P. A. T. Anxiety Scale., Instit. Person. & Abil. Testing, 1604 Coronado, Champaign, Ill., 1957.
13. CATTELL, R. B. *Personality and Motivation Structure and Measurement.* New York: World Book Co., 1957.
14. CATTELL, R. B., and BAGGALEY, A. R. The salient variable similarity index—s—for factor matching. Adv. Publ. No. 4, 1954, Lab. of Person. Assess., Univ. of Illinois, Urbana.
15. CATTELL, R. B. and COAN, R. W. Objective personality tests: factor structure at 6-7 years of age. (In Press)
16. CATTELL, R. B. and GRUEN, W. The primary personality factors in eleven year old children by objective tests. *J. Person.*, 1955, 23, 460-478.
17. CATTELL, R. B. and SCHEIER, I. H. *The Nature and Measurement of Neuroticism and Anxiety.* (In preparation)
18. CATTELL, R. B. and STICE, G. F. The Sixteen Personality Factor Questionnaire. Instit. Person. & Ability Testing, 1604 Coronado, Champaign, Ill., 1950.
19. CATTELL, R. B. and WENIG, P. Dynamic and cognitive factors controlling misperception. *J. abnorm. soc. Psych.*, 1952, 47, 797-809.
20. EYSENCK, H. J. *Dimensions of Personality.* London: Kegan Paul, Trench & Trubner, 1947.
21. EYSENCK, H. J. *The Scientific Study of Personality.* London: Kegan Paul, Trench & Trubner, 1942.
22. FREUD, S. *The Problem of Anxiety.* New York: Norton, 1936.
23. GUILFORD, J. P. Personality factors S, E and M and their measurements. *J. Psychol.*, 1936, 2, 109-128.
24. HAVERLAND, E. M. An experimental analysis by P-technique of some functionally unitary varieties of fatigue. M. A. Thesis. Urbana, Ill., Univ. of Ill., 1954.
25. HOLTZMAN, W. H., & BITTERMAN, M. E. A factorial study of adjustment to stress. *J. abn. soc. Psychol.*, 1956, 52, 179-185.
26. MAY, R. *The Meaning of Anxiety.* New York: Ronald, 1950.
27. MOWRER, O. H. Anxiety theory, as a basis for distinguishing between counseling and psychotherapy. In R. F. Berdie (ed) *Concepts and Programs of Counseling.* Minneapolis: Univ. Minn. Press, 1951.
28. ROSENTHAL, I. A factor analysis of anxiety variables. Ph.D. Thesis. Univ. of Ill. Library, Urbana, 1955.
29. THORNE, F. C. *Principles of Psychological Examining.* Brandon, Vermont: J. of Clin. Psychol., 1955.
30. TRYON, R. C. *Cluster Analysis.* Berkeley: Univ. Calif. Press, 1938.

CRITIQUE OF RECENT DEVELOPMENTS IN PERSONALITY COUNSELING THEORY

FREDERICK C. THORNE

Brandon, Vermont

Introduction

The progress of science would be expedited greatly if the originators of significant contributions would make periodic progress reports concerning their latest thinking about recent developments related to theoretical systems. Much speculative controversy could have been avoided if Freud, Adler, John B. Watson and other theorizers had issued periodic statements bringing their latest viewpoints up-to-date publicly, and particularly where important revisions of earlier viewpoints had occurred. These comments are particularly true of the field of psychotherapy where there is great need for systematically relating and validating a large number of theoretical contributions which too often have not been assimilated into the body of scientifically established facts but have become the foundations of isolated schools or systems. The purpose of this article is to present a critique of personality counseling theory bringing up-to-date our most recent thinking on the subject.

General Remarks Concerning Theorizing and System-Building

No one would voluntarily restrict creative theorizing which may supply hypothetical models for lucrative research design. There are definite dangers, however, in applied fields such as psychotherapy when theorizing and system-building far outstrip anything which has been established scientifically so that unvalidated theories are given large scale clinical application. It must be emphasized that in psychodiagnosis and psychotherapy, unproven theories and unvalidated practices are in current use far more widely than in any other clinical field. Unlike modern medicine in which schools and cults are practically extinct and where the subject matters consist largely of what has been established scientifically, in the field of psychotherapy we find a large number of competing theoretical systems and schools all being widely applied in the almost complete absence of validation and standardization.

Theory-centeredness may lead to metaphysical and dialectic confusion which actually impedes the forward progress of science. One example of such tendencies exists in the current contents of many psychoanalytic and psychiatric journals of which a formal analysis of topics and methods reveals that 80-90% of the articles are purely speculative or anecdotal in contrast with a content of about 80 or 90% of experimental-statistical studies in basic-science oriented psychological journals. Unending speculative controversy goes on concerning the theoretical implications of even chance remarks made by Freud and other "authorities". The situation is comparable to the centuries-long speculations of the greatest minds of ancient times concerning the nature of the human circulatory system which was settled forever by Harvey's simple empiric observation of the direction of blood flow in the veins by compressing and then releasing pressure on a vein on the back of the hand. The attempt to validate theories *logically* by constructing supposedly comprehensive and consistent systems is obsolete in an era when experimental-statistical validation is possible. It is regrettable that so much time is being currently spent in the attempt to secure respectibility for unvalidated clinical methods by relating them to theories which have some systematic prestige.

Psychoanalytically-Oriented Psychotherapies

Psychoanalytically-oriented practices probably have the largest current following and prestige in the fields of counseling and psychotherapy, particularly in educational institutions and clinics which are administered by psychoanalysts. Since 1945,

the psychiatry departments of several major universities have been taken over by psychoanalysts and we have reports from such installations or cities that personnel without psychoanalytic training find it difficult to obtain or retain appointments.

There is little evidence that basic science psychology is either well understood or utilized in many psychoanalytic centers. Indeed, a large portion of the time of many analysts is devoted to the training of other analysts in the same unvalidated methods, which condition may continue to exist as long as it is necessary to obtain such qualifications to gain local professional recognition.

This phenomenon of the contemporary prestige of psychoanalysis exists in spite of the fact that today, more than 50 years after the appearance of Freud's basic theories, there is still no adequate experimental-statistical validation of the system either in whole or part. While psychoanalytic formulations may receive empiric validation in individual cases, there is little acceptable evidence that they may be universally applied to all persons at all times. Without in any way disparaging the brilliance of Freud's contributions or minimizing their importance as clarifying certain classes of behavior data, there is still room for sound skepticism concerning current practices of projecting psychoanalytic interpretations on all case materials. In personal observations of psychoanalytic approaches (as well as many others), we have frequently noted therapists displaying considerable intellectual ingenuity trying to match standard theoretical interpretations to samples of behavior data but without much apparent awareness of the evident invalidity of such practices. This constitutes an example of extreme directiveness in which the therapist projects diagnoses and interpretations uncritically whether they fit or not. Our conclusion concerning psychoanalytic approaches must be that this method is essentially intuitive, anecdotal, empiric and unvalidated in its present stage of evolution since there is a dearth of experimental-statistical data. Individual applications, no matter how brilliant and productive, are at present empiric and intuitive. A great deal of the difficulty will be obviated, undoubtedly, as the phenomena uncovered by the psychoanalytic method are integrated and interpreted in terms of basic science psychology. Expediently, however, we will all go on making psychoanalytic interpretations until some better methods become available, but dangers can be minimized if greater criticality and caution are used.

Similar comments apply to projective techniques based largely on Freudian interpretations. Many clinicians go to extremes in applying projective interpretations to all behaviors ranging from global integrations down to the most incidental mannerisms. Such methods would not be too objectionable if diagnoses were properly qualified with limiting statements indicating their hypothetical and essentially unvalidated nature, but they become grossly unscientific and bordering on charlatanism when stated with absolute positivity which rules out the consideration of other possibilities. It would be wise not to forget the lessons taught by the fate of the early methods of Freudian dream interpretation. While dreams may have symbolic significance, their latent meanings only can be understood in terms of the dynamics of the individual case to which an universal dream vocabulary or symbolism can rarely be applied. It is unwise to apply projective interpretations unless there is definite evidence that projective mechanisms indeed are operating.

Client-Centered Counseling

Nondirective, client-centered methods of personality counseling have reached a stage of empiric validation where they are now widely accepted as a valuable part of our personality counseling armamentarium *within the limits of their indications and contraindications.* Utilized with clients who are motivated to understand and change themselves, and who have sufficient personality resources to work through to their own solutions once emotional blocks have been resolved, nondirective methods permit the client to reach his own solutions with maximum freedom. They are also valuable in many other clinical situations where acceptance, permissiveness and reflection of feelings facilitate the building of rapport, lead to more complete self-

expression, and encourage the development of initiative and responsibility in solving one's own problems. When such methods are utilized within their indications and contraindications, along with such other more active methods as may be indicated, and without making excessive claims for their achievements unsupported by actual research results, there can be no quarrel with their use.

Although it is too early to make a definitive statement, evidence appears to be accumulating that use of the nondirective approach is indicated more positively with "ego-negative" persons who are dissatisfied with themselves and anxious to change, and may be contraindicated with "ego-positive" persons who are satisfied with themselves and have no insight into the fact of maladjustment or the desirability of changing. Negative self-regarding attitudes, at least when not on an obsessive or psychotic basis, tend to motivate the client to face problems and work out solutions, and a minimum of therapist activity is indicated, at least until such time as direct intervention is proven more efficient. In our experience, the nondirective approach tends to fail or reach an impasse with self-satisfied clients who either see nothing wrong or who do not want to change. This includes the large group of psychoneurotic persons whose unconscious defenses may resist even massive intervention.

Much of the dissatisfaction and published criticism of nondirectivism appear to have been stimulated more by the way in which such methods have been presented or advocated rather than with the techniques themselves. If nondirectivism had been advanced as simply a promising group of new methods to be integrated with older methods according to their indications and contraindications, it probably would have been widely accepted without controversy. However, the active promotion of the client-centered principle as the basis for a new and inclusive school of theory and practice, often advocated somewhat uncritically as a new panacea suitable for all clinical situations, and even as a philosophy for the conduct of all human affairs, has stimulated critical reactions which have sought to cut the new movement down to size and integrate its contributions into existing knowledge. The efforts of nondirectivism to erect a theoretical rationale in support of its empiric achievements often have led to the making of claims which did not seem validated by actual research findings. Finally, the tendency of some nondirectivists to refer to themselves and act as if they were the exclusive proprietors of a new school whose methods would soon make obsolete all older techniques has quite understandably stimulated criticism from those who believe that all new findings must be integrated into basic science knowledge and methods.

More than ten years ago, this writer published a critique[1b] of nondirective methods of counseling and psychotherapy motivated by our concern over the apparent failure of this new school to incorporate certain standard principles of objective methodology in research and clinical approaches. A restatement of our basic criticisms in 1957 would point up five major defects of research design which seriously vitiate the conclusions which can be drawn from most of the research to date.

(a) The persistent failure of nondirectivists to recognize the importance of valid diagnosis in all scientific case study, and particularly in research design, makes it impossible to evaluate clinical claims and research results.

(b) The failure to attempt to identify all pertinent variables diagnostically has made it impossible to utilize standard research designs involving use of equated samples, rigorous identification and manipulation of dependent and independent variables, and the objectification of all factors and steps known to be involved.

(c) The failure to control, or otherwise partial out, the effects of other therapeutic factors (such as suggestion, catharsis, desensitization, etc.), which are known to be operating, makes it impossible to draw any valid conclusions as to what is actually occurring in nondirective therapy.

(d) The failure to check conclusions drawn from phenomenological data obtained from the client (or inferred by the therapist) with actual objective measurements of known factors or status may result in serious misinterpretations of the realities of the therapeutic situation.

(e) Even if positive therapeutic results are demonstrable using nondirective methods, it still remains to be proven that such effects could not have been achieved just as well or even more expeditiously with other methods.

In conclusion, then, the failure to rigorously identify, diagnose and control all factors known to be operating in psychotherapy has seriously invalidated much of recent research on nondirective methods. We are not told whether the subjects are normal or abnormal, the nature and pathogenicity in terms of actual case history findings of the factors being therapized, and we are not convinced that the postulated factors actually explain the results obtained. In other words, most of the factors which we need to know about have been left undifferentiated and uncontrolled. In effect, it appears that nondirectivism postulates one standard diagnosis for all clients, namely that emotional factors blocking growth are the most important cause of all disorder, hence diagnosis is not necessary and only one standard method (nondirectivism) is indicated.

The most prominent success of client-centered methodological research has been in objectively demonstrating the actual operations which take place in this method. Nondirectivism has made a very great contribution in developing passive methods of counseling emphasizing acceptance, permissiveness, reflection and clarification of feelings, and placing the responsibility upon the client for working out his own problem solutions. Conversely, the dangers of crude directiveness, counselor-centeredness, and the invalid projection of interpretations upon the client's behavior are all better understood as a result of insistence on the client-centered principle. The issue still remains, however, concerning indications and contraindications for use of nondirective methods with different types of clients, pathological processes, degrees of pathogenicity and other factors determining outcomes.

Proponents of nondirectivism have been less successful in trying to construct a systematic theoretical rationale for their viewpoint. It appears that Rogers' earliest contributions were developed by intuitive-empiric methods, interpreted at first[4] in terms of somewhat mystical concepts of self-realization and growth, and more recently developed in terms of a thoroughgoing phenomenological approach[5]. While such preoccupation with the self-concept and other phenomenological data as experienced by the subject himself and understood empathically by the therapist is very intriguing and commendable in itself, considerable criticism may be directed toward the attempt to construct a complete system on such a limited approach. Even more caution should be invoked in evaluating claims by nondirectivists for unique processes and results from their methods, particularly when it is known that all therapeutic methods have their successes. Before we can ascribe much validity to claims made, for example, from changes on Q sorts given pre- and post-client-centered therapy, it will be necessary to obtain comparable Q sorts from clients treated with Coueism, Christian Science, Dianetics, pyschoanalysis and many types of directive methods (all of which have very enthusiastic clients and practitioners). It may be that certain standard factors underly all these methods and that specific philosophical rationales are superfluous. Snyder recognizes this dilemma when he states that "Perhaps it doesn't make very much difference which philosophies guide a therapist's work".[9] While it is interesting to attempt to trace the philosophical antecedents of recent theoretical developments as when Walker[16] perceives Freud as the intellectual disciple of Augustine, and Rogers the intellectual descendent of Jesus, Confucius or Rousseau, there are dangers in implying that psychoanalysis or nondirectivism have some face validity from being related to such respected authorities. Similar comments apply to the currently popular practice of using laudatory or discriminatory designations to categorize accepted or rejected concepts as when Snyder[9] categorizes neofundamentalism, scholasticism, behaviorism, psychobiology, learning theory and directive counseling as being "authoritarian", while nondirectivism is identified with "self-actualism" or "self-determinism", idealism, democratic government, humanism, progressive education, hormic psychology and phenomenology. Such a dichotomy is purely arbitrary and probably artifactual, and even worse are the self-reflexive connotations of the term "authoritarian" as interpreted against the recent associative contexts of Fascism, Communism, regimentation and exploitation as contrasted with other possible connotations involving such

positive contexts as love, protection, paternalism, guidance, help and wise counseling. The use of such self-reflexive terms in purportedly scientific communications is very undesirable.

While we would defend to the end the right of Carl R. Rogers[6] to expound his own philosophical queries and experiences in relation to problems of science and phenomenology, we must insist that the evidential value of such observations must be rigorously evaluated by scientific methods before being accepted as anything but expressions of personal opinion and faith. The quandaries of Rogers in relation to "Persons or Science"[5] appear to be rationalizations of the deficiencies of his ideological approach. Rogers' rejection of diagnosis as a basic method of research and of clinical practice has seriously limited the scope of his whole approach, and his failure to modify his position, to publicly indicate the limitations and contraindications of nondirectivism, or to discuss their integration with older methods, may actually inhibit the acceptance of the whole movement. Snyder appears to admit this in his somewhat despairing comment of a "pox on both sides" as he concludes that "as long as Rogers cares and the Freudians care, both seem to be meeting the needs of clients."[9]

Actually, the defects of the research design under which most nondirective research has been done could be avoided by the simple procedure of having some independent authority select and thoroughly study samples of cases before submitting them for nondirective treatment. This device would simultaneously objectify the nature of the case materials while at the same time free the client-centered therapist from introducing any sort of diagnostic procedure during his case handling. After the termination of therapy, the independent authority could again study the cases for the presence and nature of any changes occurring from therapy. Such results could then be compared with other samples treated by other standard methods, and the data subjected to factor analysis to discover any common factors underlying all methods.

Eclectic Counseling

The eclectic method of selecting appropriate theories and methods to fit the indications and contraindications of each individual case has long been the standard procedure in the older clinical sciences such as medicine. In personal reaction to the seeming one-sidedness of nondirective and psychoanalytic approaches, publication of a series of articles on directive therapy was begun in 1945 and later expanded into a systematic textbook.[12] Our adoption of the term *directive* was intended only to emphasize the existence of a large number of therapeutic methods (including nondirectivism) within the larger framework of all that was known about influencing personality. Our use of this term was *not* intended to imply any preference for any special group of methods or philosophical viewpoint, and we regret that many readers immediately projected many unintended meanings which now we do not wish to be identified with. The term *eclectic* would have been more accurate semantically and, by 1948, we specifically attempted to limit the term *directive* to the systematic attempt to use all methods of therapy eclectically based on a formal plan for the identification and modification of etiologic factors in maladjustment.[11] Now, we prefer to abandon all terms such as *directive*, *nondirective*, *psychoanalytic*, etc., as tending to create artificial continua and semantic artifacts, and instead to substitute operational descriptions and nomenclature for the dynamics of any method. The term *eclectic* has value only to differentiate an approach which refuses to subscribe to any one operational viewpoint or method, and becomes entirely objectionable if given self-reflexive connotations such as "authoritarian".

The eclectic approach to scientific case handling basically depends upon adequate diagnosis as the cornerstone of all case handling in order to identify etiologic causes and apply appropriate therapeutic methods according to their indications and contraindications. Currently, the most important theoretical issue in personality counseling and, indeed, all psychotherapy, is the matter of the relation of diagnosis

to valid therapy. Far from attempting to establish any new school or system based on "directive" methods, our sole purpose has been to emphasize the importance of diagnosis as the cornerstone underlying the rationale for knowing the indications and contraindications of *all* known methods utilized without systematic bias of any kind. Frankly, we have been very much disturbed over the past few years by the discovery that many teachers and clinicians have entirely misunderstood our fundemental theoretical position which stresses a thorough-going basic-science-oriented eclecticism by erroneously identifying us with directivism as a method or end in itself. Our original selection of the term "directive psychotherapy" was intended only to emphasize that an important group of methods exists in contrast to nondirectivism, and it was not intended to identify a new school or system. Now the need for emphasizing such a difference has long since passed, and we would prefer to use the term directive only to indicate one end of a continuum of passivity *vs.* activity in the therapist. In our own practice, we try to be passive or active according to the indications of the situation rather than following any ideological bias.

One of the criticisms advanced, particularly by nondirectivists, against eclecticism in its present status of development, relates to the voluminous research which has been accumulated by special schools in contrast with the meagre or nonexistent research output achieved by the exponents of eclecticism. Such criticisms have been directed against eclectic or directive counseling as where Shoben says that "Thorne is unable to escape an authoritarian tone that could not be challenged in itself if it were buttressed either by experimental evidence or convincing case material"[8, p. 43], and which comment has appeared repeatedly in many forms. Such a criticism appears to reflect a lack of understanding concerning the differing problems and responsibilities placed upon those who accept eclecticism in contrast with those who support more limited special methods. Our original aim in attempting to compile a unified systematic approach to diagnosis and therapy was to attempt to gather in one place all that was scientifically valid and fairly well accepted from all sources or schools, both clinical and experimental-statistical, eclectically. We have never made any claims for any of these methods beyond what had been advanced in the literature or appeared valid in terms of personal clinical experience. We submit that this was all that could be accomplished at time and place, and we make no special claims for any superiority of any of these methods beyond what has been empirically established. Our role was simply to attempt an integration of all that seemed to be accepted in terms of the accumulated clinical experience of our generation and without favor or prejudice for any method beyond the known indications and contraindications as currently understood. This task of attempting to integrate, in one internally-consistent presentation, the accumulated experience of clinical science is in itself so gigantic as to preclude any personal research attack upon the vast number of problems and methods, and certainly beyond our personal resources which have been invested maximally in the attempt to integrate existing research interpretively. We can only state, then, that our conclusions represent only the concentrate of human knowledge at time and place as we have interpreted it. We feel that there is a definite need not only for researchers who can provide valid data but also for global interpreters who must attempt to relate the many disparate findings which are in themselves more or less meaningless until oriented in some master frame of reference.

Any deficiencies of eclecticism or directivism simply reflect either (a) the relatively primitive state of human knowledge concerning these methods, or (b) the inadequacies of breadth of viewpoint among those who attempt an integrative interpretation. To the degree which eclecticism is able to integrate *all* operational methods and findings available at time and place, it appears to us that it must represent the last word concerning what we can validly understand and apply in practice. We may accept, as a basic premise, the axiom that clinical methods are simply individual applications (within the limits of validity of clinical judgment) of basic science experimental-statistical methods and data. We seek to protect the validity of our judgments and conclusions by rigorously seeking to discriminate cause-effect relationships

by the most advanced diagnostic methods. We submit that this is all which can be done for time and place within any clinical orientation. Eclectic methodology is so diverse and complex that a general advance on all fronts simultaneously can come only very gradually and certainly beyond the scope of any one person or generation to advance them very markedly. In spite of this admitted deficiency based on the primitive status of existing knowledge, we can be assured of some degree of face, content, empirical and factorial validity by constantly seeking to relate our practices to the most advanced basic science thinking of time and place.

In contrast, the proponents of special methods have a particular obligation to demonstrate that new ingredients have been added over and above existing knowledge. Considering the impressive basic science advances of the last century, both by science in general and by special schools, the burden of the proof rests upon proponents of new schools to show where and how they have advanced human knowledge beyond merely introducing new terms or rehashing the known. Modern medicine for example, has reached a status of solid basic science achievement where we can expect new discoveries in special areas by experimental-statistical methods, but we do not anticipate any revolutionary philosophical advances which will change our basic methods of scientific discovery or thinking. Science supercedes schools of philosophy or metaphysics as soon as it accumulates sufficient objective knowledge in any field to provide valid bases for clinical applications. We believe that clinical psychological science has already achieved such a solid basic science foundation of proven facts as to make obsolete the need for metaphysical theorizing and special schools. If new advances are made, they must be integrated into the broad foundation of scientific knowledge rather than being offered as revolutionary panaceas to replace all that is eclectically known.

Eclecticism in clinical practice is always held back in situations where students are trained in only one method, and are either oblivious of other methods or insecure in their application. Until such time as students are provided training experiences with all known operational methods of diagnosis and therapy, it is evident that their capacity or competence in applying all known methods eclectically will be limited. The need to build up one's confidence clinically by becoming a "master" in some specific school, such as nondirectivism or psychoanalysis, disappears as personal competence with all known methods broadens and one is sustained by the knowledge that one knows all there is to know for time and place, *i.e.*, broad scientific knowledge and clinical competence reduce the need for speculative theorizing.

Disciplinary Counseling

Williamson and Foley[17] have presented a detailed reporting of the use of "counseling" in the disciplinary management of student conduct problems as developed since 1941 in the office of the dean of students at the University of Minnesota. Although the authors evidently are aware of the classically *therapeutic* basic role of personality counseling, and repeatedly emphasize the rehabilitational and reeducative objectives of their program, it becomes immediately apparent that the term "management" would be a much more realistic designation of their methods of case handling than is the term "counseling". Although in defining their concept of the role of counseling in student discipline, the authors are careful to structure their activities as not being authoritarian, regulatory or punitive, an analysis of their outline of case handling methods and the actual transcripts of individual cases indicates that their activities are conducted within a plainly authoritarian judgmental framework which can hardly be perceived as anything else by the students. In spite of what Williamson and Foley theorize the situation to involve, the critical element in categorizing the nature of the situation must be in terms of what the subjects of the counseling perceive it to be. In discussing their procedures of disciplinary counseling, their schematic outline[7, p. 61] includes such steps as (a) making of charges against the student, (b) compulsory interviewing for counseling purposes, (c) reporting and review by a disciplinary committee in a face-to-face situation with the

student, and (d) disciplinary action by the committee involving the imposition of further regulatory actions upon the student. Whatever the professed intentions of the disciplinary committee, the execution of its functions in such a framework cannot avoid the connotation of a judgmental, regulatory and, perhaps, punitive orientation to case handling as the whole situation is perceived by the student from the implications of what goes on. That many students may, indeed, perceive the situation as basically judgmental and punitive is clearly evident in the case of Milton Reed [7, p. 229] where not only the accused but also two friends openly profess their fears that unreasonable judgments and/or punishments are going to be administered.

In our opinion, no method of case handling deserves to be designated as counseling unless *healing* is the basic and primary objective beyond which all other considerations are secondary. In order to retain its basically therapeutic functions, the counseling situation must be thoroughly safeguarded to preserve its purpose of being the one place (in addition to psychiatry) where a person may frankly discuss his problems in a completely healing atmosphere unthreatened by any question of judgment or punishment. We are in agreement with Rogers [4] in his contention that any other structuring of the situation destroys the therapeutic context.

Implicit in the approach of Williamson and Foley is the questionable assumption that misbehavior is classifiable into voluntary and involuntary types for which discipline and therapy are the solutions respectively. While the authors recognize that immaturity and personality deviations may interact with unhealthy environmental forces to produce misconduct, the installation of such complicated judicial machinery and practices only succeeds in advancing the implication that we are really dealing with deliberate misbehavior, bordering on criminality, and which can be handled best by punitive procedures progressively applied until the miscreant sees the light and voluntarily controls himself better. We must insist that such methods of case handling are more properly designated as management and do not qualify as counseling because the basic operational framework is not primarily therapeutic but rather is regulatory and administrative. This is not to imply that counseling cannot be carried on within disciplinary or administrative auspices where the healing motive is primary and disposition is secondary. *We would prefer to base all counseling on the postulate that all misbehavior reflects immaturity, personality deviation or personality reactions to situations, reflecting various degrees of lack of control, and arising etiologically in the domain of psychopathology.* From such a viewpoint, all disciplinary problems should be handled within a completely therapeutic frame of reference, only being referred for regulatory action in the event of the breakdown of therapeutic progress and poor prognosis. The authorities could take notice of misconduct by anonymous comments on the situation, clearly indicating its position that such behavior must be the result of unhealthy thinking, and further indicating that gross regulatory methods would be enforced only in the event of the breakdown of rehabilitation.

There is serious question whether disciplinary counseling as practiced in the case of Janet Larson [7, p. 261] may be actually untherapeutic and dangerous to the health of the subjects concerned. We can think of few situations more traumatic and guilt-productive than for students to be detected in the act of sex intercourse in the bushes by the university watchman and such detection is usually punishment enough. Psychiatrically, there is great danger of the girl (at least) developing incapacitating mortification and guilt reactions which may lead to life-long neuroticism and frigidity. When this student was interrogated by F. L. Crane in the presence of John D. Foley on July 22, 1945 in great detail concerning what took place, the circumstances of the interview indicate that it was conducted somewhat like a third degree solely for the purpose of securing a signed confession and with nothing therapeutic taking place. Further, on July 23, the accused Janet Larson and William Burton were called before the faculty disciplinary committee and interrogated together about their marriage intentions, guilt reactions and penitent attitudes. The trend of the questioning seemed to hint that the university would be satisfied if the couple married (shot gun influence?) but it actually did not accomplish this purpose,

perhaps because Burton was led to admit previous intercourse before Miss Larson and also did not seem completely enthusiastic about marriage. The most therapeutic action taken was the decision of the disciplinary committee to put the students on probation while continuing at the university. It is interesting to note that on May 11, 1946, counselor G. E. K. spontaneously notes that Miss Larson no longer has "a vivacious, energetic enthusiasm" in her manner but has changed to "quiet winsomeness", which might be interpreted as reflecting deep anxiety or guilt. In any case, a perusal of the case record reveals no evidence of therapeutic counseling but only the administrative details of handling student misbehavior.

An analysis of the entire 15 cases cited as examples of disciplinary counseling reveals that the counselor activity seems to consist in acting as the representative of university authority in identifying alleged misbehavior, identifying the students involved, obtaining signed confessions before witnesses, conducting judicial hearings, and later advising the student how to behave in order to stay out of trouble again. While some secondary therapeutic results may be achieved where leniency, tolerance and patience may operate to keep the student in the university until he can learn to handle objective behaviors more adaptively, there is little evidence either that the primary emphasis is therapeutic or that modern counseling methods have been used under maximally effective conditions to secure deep personality changes.

Different Operational Approaches

Order may be brought into the current confusion of competing psychotherapies by reducing or translating each approach to its lowest operational common denominator. As pointed out elsewhere[13, p. 138], behavior analysis may be attempted from the operational viewpoints of instinct theory, perceptual field theory, learning theory, motivation theory or the psychology of the emotions. Each diverse approach utilizes its own operational levels of personality study and achieves its own distinctive results which may reflect determination by general, group or specific factors. Clinicians trained primarily in one operational approach tend to have trouble comprehending the methods and results from other operational approaches. The advantage of a thoroughgoing operational approach is that it permits the study of raw behavior data directly and with the formulation of diagnostic impressions *inductively* to fit the individual case with the progressive interpretation of data. The problem is to construct a theory to fit the case inductively rather than trying to apply theories deductively. Operationism renders unnecessary undue speculation and theorizing if carried out in sufficient detail to uncover the major etiologic factors in each case.

The fact that many methods appear to produce results can only have two possible explanations. Either similar dynamic factors are operating in diverse methods, or there is more than one method for achieving the same specific goals. For example, as pointed out by Snyder[8, p. 92], perhaps a theory of psychotherapy does little more than to inspire the therapist with confidence which is transmitted to the client and becomes the basis of a relationship which shows the client that "somebody" cares. Or perhaps suggestion effects and emotional release are enough to help most clients. Or perhaps merely doing something until time and nature heal is enough. Such important questions can only be solved with a research design complex enough to identify and control all relevant variables. In the meantime, symposia on perceptual field theory or learning theory in relation to psychotherapy may be very valuable in casting light on what these various operational approaches may contribute but it should not be expected that any one operational approach can provide the complete answer. Combs[2] may be correct in concluding that traditional learning theories offer little towards the improvement of counseling (that is, beyond the general recognition that all therapy involves learning according to well-known psychological laws), but we feel that he is overoptimistic in assuming that all behavior is a function of the field of perception at the moment of behaving so that the process of counseling becomes one of altering perceptions in desired directions. Shaw[6] apparently takes a much more valid position (in agreement with our eclectic view-

point) that varied conditions of learning in persons may produce a variety of etiologic factors operating on many levels of personality integration. We agree with Shoben[7] that contemporary learning theories are most interested in *principles* of learning while the clinician must attempt, as we have also previously pointed out, to (a) establish suitable conditions so that learning can take place, and then (b) to actually introduce new conditionings in more healthy directions.[14, p. 278] Our own detailed viewpoint is that personality integration *may* be organized by many levels of factors, which may be perceptual, learned, motivational, affective or instinctive, operating at various levels of generality or specificity, and each requiring their own particular operational methods for diagnosis or treatment.[15, p. 14-15] It will probably require complex experiments of factor analytic design to differentiate prototypical patterns of syndrome organization.

Semantic Problems

A most urgent problem in the whole field of diagnosis and therapy concerns the early adoption of a uniform and universal system of nomenclature so that the coordinates and boundaries of psychological space may become standardized. Every few months some authority launches a new theory replete with its own distinctive terminology which usually is not related to the coordinates of accepted knowledge. Two recent examples of such system-building with difficult new terminologies are Allport's event-structure theory[1] and J. G. Miller's general behavior systems theory[3], both of which will require long intensive study to become proficient with and simply to master the terms with any facility. And even after the new terminology is mastered, there still remains the problem of relating the new contributions to existing knowledge to discover whether or not the new formulation makes any distinctive contribution. There is great need for the profession to adopt a standard nomenclature with a universal index of factors to be measured. The most promising beginnings in this direction have been made by the factor analysts who are working towards standard methods for labelling and mapping new factors in their space-time dimensions, and we have personally adopted standard terminology in our approach to psychological examining. Semantic chaos will result from the multiplication of new theories and systems unless standard nomenclature is uniformly adopted to relate new developments to existing coordinates. In our own editorial work, we uniformly reject papers introducing new and esoteric terminology unless it can be clearly shown that such refinements are needed to describe genuinely new developments. As matters now stand, students entering the field must spend entirely unreasonable amounts of time trying to master the babel of conflicting terminologies.

Predictions for the Future

The most urgent theoretical question in personality counseling and, indeed, all psychotherapy, is the matter of the relation of diagnosis to valid therapy. Currently, it is fashionable in some circles to ignore problems of diagnosis or to categorize such interests as authoritarian.[8, p. 43] It is a basic postulate of all scientific method to attempt to discover the nature and etiology of all variables determining a natural occurrence, *i.e.* to identify causes and effects. Without in any way attempting to defend the inadequacies of current diagnostic methods and systems, we cannot see how any progress can be made in understanding causation or rationally manipulating determinants unless a diagnosis is made of all antecedents leading to consequences. Our opinion is that advances in psychotherapy must depend on more valid diagnosis and to this end we have evolved a systematic method[15] for evaluating all factors known to organize personality integration. We emphatically reject the hypothesis that psychotherapy is in any way different from other clinical science in the sense of it being possible to dispense with diagnosis in scientific case handling. This has nothing to do with any philosophic bias but is a tenet of scientific method.

Within this general basic science orientation, we may predict future developments in the issues which are now so controversial. We predict that the current

heavy investment of time and energy in speculative theorizing and system-building will be supplanted by a consistent effort to relate all knowledge and new developments in terms of a standard nomenclature and universal index of factors and results. We predict that exclusive preoccupation with special operational approaches such as perception or learning theory will be replaced by a broader eclectic utilization of all available operational approaches according to the indications and contraindications of time and place. We predict that nondirectivism (and all other special therapies) will be demoted from their current status as mutually exclusive systems to take their proper places as limited technical advances integrated into a broad diagnostic-therapeutic armamentarium applied eclectically, without special bias or preference for specific tools. We predict that much of psychoanalytic speculation and dialectics will be abandoned, either because ultimately proven to be unvalidatable or more parsimoniously explained in terms of basic science concepts, and leaving a residual of fact which can be integrated into basic science knowledge in the proper places. We predict that all clinicians will accept ultimately the absolute essentiality of diagnostic thinking and will cease to seek short-cuts motivated by current deficiencies. In other words, we reaffirm our belief in the ultimate victory of scientific methodology in proving its capability of dealing with all psychological problems without recourse to less rigorous philosophizing and speculation.

We cannot escape the fact that all natural phenomena reflect the operation of laws which determine causes and effects. If we are to understand and manipulate these causes and effects with maximum efficiency, then we must recognize the various categories of action imperatives involved in the concept of "ought". Elsewhere [14], we have explored the implications of the psychology of "ought" for social planning in general, and individual efficiency in particular, with special reference to what can be accomplished in diagnosis and therapy. We insist that all this can be accomplished within the orientation of eclecticism. We hope that adherents of various special positions will cease projecting various characterizations upon each other and will work in common to discover and control the actual etiologic factors which must be operating.

References

1. Allport, F. H. *Theories of Perception and the Concept of Structure.* New York: Wiley, 1955.
2. Combs, A. W. Counseling as a learning process. *J. counsel. Psychol.*, 1954, 1, 31-36.
3. Miller, J. G. General behavior systems theory and summary. *J. counsel. Psychol.*, 1956, 3, 120-124.
4. Rogers, C. R. *Counseling and Psychotherapy.* Boston: Houghton-Mifflin, 1942.
5. Rogers, C. R. Persons or Science? A Philosophical Question. *Amer. Psychol.*, 1955, 10, 267-278.
6. Shaw, F. J. Counseling from the standpoint of an "*Interactive Conceptualist*". *J. counsel. Psychol.*, 1954, 1, 36-42.
7. Shoben, E. J., Jr. Introduction: Behavior Theories and a Counseling Case. *J. counsel. Psychol.*, 1956, 3, 107-108.
8. Shoben, E. J., Jr. The mosaic for psychotherapy. *Contemporary Psychol.*, 1956, 1, 42-44.
9. Snyder, W. U. Comment. *J. counsel. Psychol.*, 1956, 3, 91-92.
10. Thorne, F. C. A critique of nondirective methods of psychotherapy. *J. abn. soc. Psychol.*, 1944, 39, 459-470.
11. Thorne, F. C. Principles of directive counseling and psychotherapy. *Amer. Psychologist*, 1948, 3, 160-166.
12. Thorne, F. C. *Principles of Personality Counseling.* Brandon, Vt.: Journal of Clinical Psychology, 1950.
13. Thorne, F. C. Operational concepts in clinical psychology and medicine. *J. clin. Psychol.*, 1953, 9, 137-142.
14. Thorne, F. C. Directive psychotherapy: Theory, practice and social implications. *J. clin. Psychol.*, 1953, 9, 267-280.
15. Thorne, F. C. *Principles of Psychological Examining.* Brandon, Vt.: Journal of Clinical Psychology, 1955.
16. Walker, D. E. Carl Rogers and the Nature of Man. *J. counsel. Psychol.*, 1956, 3, 89-91.
17. Williamson, E. G. and Foley, J. D. *Counseling and Discipline.* New York: McGraw-Hill, 1949.

THE CONSTRUCT VALIDITY OF THE SIXTEEN PERSONALITY FACTORS TEST

SAMUEL KARSON[1] AND KENNETH BRYNER POOL[2]

1100th U. S. A. F. Hospital
Bolling A. F. B., Washington, D. C.

School of Aviation Medicine, USAF
Randolph A. F. B., Texas

PROBLEM

Cronbach and Meehl [5] have examined the construct of validity in psychological testing and discussed four types of validation—namely, predictive validity, concurrent validity, content validity and construct validity. The first two kinds of validity can be distinguished from the last two in terms of criterion-oriented validation procedures. The present paper is concerned with construct validation which "is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned and must use indirect measures." [5, p. 282] As Cronbach and Meehl have indicated, in studies of construct validity the trait or quality underlying the test is of central importance, rather than either the test behavior or the scores on the criteria. In their paper they mention [5, p. 300] that many types of evidence are relevant to construct validity, including content validity, interitem correlations, intertest correlations, test "criterion" correlations, studies of stability over time, and stability under experimental intervention.

The present study is an attempt to evaluate certain aspects of the construct validity of the Sixteen Personality Factors Test principally by means of correlating its factors with the better-known scales of the MMPI. Specifically, the following questions were investigated: What are the relationships between MMPI scale scores and 16 P. F. factor scores? Do the two forms of the 16 P. F. Test correlate highly enough to be used as equivalent forms as they were intended [3], *i.e.*, what are the coefficients of equivalence of the individual factors? Further, what is the relationship between scores on factor B (general intelligence) and scores on individual tests of intelligence such as the Wechsler-Bellevue Intelligence Scale Form I (W-B I) and the Wechsler Adult Intelligence Scale (WAIS)? Finally, an attempt was made to determine whether experienced clinicians are capable of predicting which factors from the 16 P. F. would be significantly correlated with each of the MMPI scales. The design of this study should make apparent which areas of personality the 16 P. F. Test and the MMPI measure in common as well as those areas which are idiosyncratic to each.

PROCEDURE

The sample consisted of 71 USAF officers on flying status who were referred to the School of Aviation Medicine for medical and psychological evaluation. The average age was 33, with a range of 20 to 39 years, and the mean number of years of education was 13, with a range of 12 to 16 years. The mean Full Scale IQ on the Wechsler Adult Intelligence Scale was 121.98 with a standard deviation of 6.50. The patients were referred for evaluation between August, 1955, and July, 1956, and were individually administered the group form of the MMPI which contains 566 items by the Department of Neuropsychiatry, and Forms A and B of the 16 P. F. Test by the Department of Clinical Psychology. Usually no more than 3 days elapsed between the administration of these two tests. The two forms of the 16 P. F. were alternated so that one half of the sample had Form A first, while the other half had Form B first. The 16 P. F. was given as a part of the routine diagnostic psychological testing battery. Both forms of the test were usually administered on the same

[1] The data for the study were collected and the statistics computed while both authors were at the School of Aviation Medicine, USAF.

[2] Appreciation is expressed to Lts. John K. Sussenberger and Gordon Raynor for their aid in the administration and scoring of the tests, and to Dr. S. B. Sells for suggesting the comparison between the 16 P. F. Test and the MMPI.

day with the form taken first given in the morning, and its counterpart in the afternoon. The sample employed in this research is believed to be typical of the flying personnel who are referred for consultation in the United States Air Force.

The tests were scored according to the directions given in their respective manuals[3, 4, 13]. The raw scores for Forms A and B were converted into standard scores by using the sample of male college students and high school seniors whose average age was 21 as the standardization group. Pearson coefficients of correlation within and between the two tests were computed, using the raw scores of the 14 MMPI scales and the standard scores of the 16 P. F. Test.

Results

In every case high scores on the MMPI were correlated with high scores on the 16 P. F. Test. It should be noted that on some of the factors of the 16 P. F. a high score does not have a negative connotation as it always does on the MMPI. The matrix of intercorrelations is presented in Table 1. The large number of significant correlations obtained between the 16 P. F. factors and MMPI scales (87 out of 244 or 39%) demonstrates there is a common area of overlap between these tests of personality. As might be expected, fewer significant correlations occur among the 16 P. F. factors (39 out of a possible 120 or 32%) than among the MMPI scales (61 out of a possible 91 or 67%). These results are a function of the factor analytic construction of the 16 P. F. as compared with the strictly empirical method used in the construction of the MMPI.

Looking now at the significant relationships which the MMPI and 16 P. F. Test (Forms A and B combined) share with one another, it can be observed from Table 1 that apparently the latter test adequately taps all of the areas of personality measured by the MMPI with the exception of the Pa and ? scales. Each of the other 12 MMPI scales is significantly correlated with a minimum of four factors from the 16 P. F., even though the MMPI is the larger of the two tests by approximately 30%. When the MMPI scales are matched with the factors from the 16 P. F. Test, 7 of the 16 factors are correlated significantly with 3 or fewer scales of the MMPI—namely, factors A, B, E, G, N, Q_1 and Q_3. Apparently the dimensions of personality measured by these factors are not well represented in the MMPI. These results appear to demonstrate certain advantages in economy of test length, as well as in breadth, for the factor analytically constructed 16 P. F. Test over the empirically constructed MMPI.

The direction of certain of the specific relationships observed in Table 1 are notable. Curiously enough, L scores on the MMPI correlate positively with scores on factor C (ego strength), and negatively with factors O (anxious insecurity) and Q_4 (nervous tension). Further, Hy correlates positively with factor N (sophistication) and Mf correlates positively with factor B (general intelligence), negatively with factor F (surgency) and positively with factor Q_3 (independent self-sufficiency), while Ma correlates positively with factor L (suspecting, jealous). The other 79 significant correlations between the two tests reported in Table 1 all appear to be in the direction to be expected on the basis of the descriptions presented in the respective manuals and from clinical experience as well.

Coefficients of equivalence for the 16 factors of each form with order of administration controlled, are reported in Table 2. With Form A administered first, 9 of the coefficients fall below .50, while only 2 of the factors (H and Q_2) exceed .70. While with Form B administered first, 11 of the reported coefficients fall below .50, and only one factor (H) exceeds .70. For the total sample, average Pearson coefficients of correlation between the two forms indicate that they are not equivalent enough to be considered satisfactory alternate forms. Even when coefficients of equivalence are computed with the Spearman-Brown prophecy formula for determining the effects of doubling the length of the test[7], 6 of the reported coefficients still fall below .50, and only 5 of the factors (A, H, O, Q_2 and Q_4) exceed .70. These results suggest that for accurate individual diagnosis in clinical work, neither of the test forms

TABLE 1. INTERCORRELATIONS OF FACTOR SCORES ON THE SIXTEEN P. F. TEST, FORMS A AND B COMBINED, AND MMPI SCALES (N = 71)

Factors		A	B	C	E	F	G	H	I	L	M	N	O	Q_1	Q_2	Q_3	Q_4	K	F	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	L	?
Cattell	A		02	12	13	42*	-10	45*	-12	-07	-08	12	-07	01	-27*	-08	11	13	04	-03	-11	20	-09	-14	08	-06	-04	00	-33*	05	-02
	B			-12	08	00	-08	-11	02	21	18	-13	21	-15	02	-15	20	-14	09	02	-02	-07	06	29*	22	13	11	-05	18	-24*	-02
	C				22	30*	-17	41*	-38*	-59*	-57*	03	-71*	-40*	-13	43*	-64*	58*	-48*	-54*	-36*	-18	-48*	-48*	-17	-63*	-66*	-30*	-39*	34*	-03
	E					41*	-03	49*	-43*	00	-07	13	-32*	12	-01	-13	-08	23	-07	-04	-11	05	02	-35*	22	-21	-24*	10	-27*	06	13
	F						-20	66*	-23	-14	-14	12	-28*	-02	-16	-19	-13	08	-01	-14	-26*	05	01	-25*	-04	-14	-11	24*	-48*	06	-04
	G							08	16	23	03	-07	00	07	26*	17	04	-18	-04	08	14	-06	19	09	03	08	03	13	04	07	-06
	H								-26*	-23	-27*	18	-52*	-17	-36*	05	-28*	27*	-15	-26	-24*	05	00	-35*	08	-31*	-30*	19	-69*	19	-07
	I									20	45*	-23	40*	04	16	-09	24*	-30*	30*	31*	21	16	23	54*	02	34*	41*	15	23	03	00
	L										52*	00	46*	22	22	-18	51*	-63*	33*	29*	13	-05	45*	33*	05	55*	50*	30*	41*	-25*	21
	M											-13	56*	31*	29*	-36*	53*	-48*	46*	38*	17	16	29*	47*	01	48*	48*	17	32*	-15	16
	N												-18	11	-01	-03	-12	01	04	07	-02	25*	01	-16	04	-08	-03	06	-08	-12	00
	O													21	23	-41*	76*	-59*	49*	51*	45*	26*	39*	63*	15	77*	68*	21	54*	-26*	21
	Q_1														24*	-16	17	-10	24*	22	22	30*	19	24*	15	17	21	11	11	-04	10
	Q_2															01	09	-22	03	05	01	-03	-02	27*	-15	07	07	03	32*	-05	11
	Q_3																-42*	35*	-40*	-32*	-21	-13	-27*	-17	-21	-41*	-38*	-27*	-11	-23	-12
	Q_4																	-63*	51*	62*	52*	25*	38*	48*	23	75*	64*	14	45*	-38*	26*
MMPI	K																		-52*	-45*	-27*	04	-52*	-51*	-06	-76*	-69*	-36*	-52*	32*	-30*
	F																			64*	40*	39*	54*	44*	34*	65*	71*	39*	23	-19	18
	Hs																				66*	62*	32*	38*	26*	60*	56*	11	35*	-15	31*
	D																				..	49*	29*	33*	29*	55*	40*	-21	44*	-17	29*
	Hy																						16	16	24*	27*	24*	-04	-04	00	14
	Pd																							36*	45*	66*	66*	54*	11	-34*	15
	Mf																								22	61*	58*	16	46*	-21	20
	Pa																									36*	36*	18	-04	-20	11
	Pt																										89*	35*	53*	-41*	29*
	Sc																											45*	40*	-37*	15
	Ma																												-23	-16	-09
	Si																													-24*	33*
	L																														04
	?																														

*Denotes significance. For significance at the 1% and 5% levels of confidence, Pearson r's must equal .31 and .24.

TABLE 2. A COMPARISON OF THE MEANS, STANDARD DEVIATIONS AND COEFFICIENTS OF EQUIVALENCE OF THE CATTELL SIXTEEN P. F. TEST FORMS A AND B

	Form A Administered First (N = 30)					Form B Administered First (N = 30)					Total Group (N = 60)					
	Form A		Form B			Form A		Form B			Form A		Form B			
Factor	Mean	S.D.	Mean	S.D.	Pearson r*	Mean	S.D.	Mean	S.D.	Pearson r*	Mean	S.D.	Mean	S.D.	Pearson r**	Estimated r***
A	5.23	1.52	5.52	1.88	.56	5.06	1.63	4.32	1.89	.57	5.15	1.57	4.92	1.89	.56	.72
B	4.87	1.93	6.03	1.73	.21	4.61	1.90	5.87	2.09	.34	4.74	1.91	5.95	1.91	.27	.33
C	5.55	2.08	7.32	1.47	.45	5.71	2.36	7.13	1.68	.19	5.63	2.22	7.23	1.57	.33	.50
E	5.58	1.77	4.97	1.75	.42	5.52	1.52	5.10	1.57	.18	5.55	1.65	5.03	1.66	.31	.38
F	4.71	1.46	4.71	1.94	.57	4.03	1.38	4.19	1.86	.49	4.37	1.42	4.45	1.90	.53	.69
G	5.23	1.83	5.81	1.99	.37	5.65	1.75	6.19	1.99	.60	5.44	1.79	6.00	1.99	.49	.66
H	5.61	2.10	6.03	1.76	.84	6.23	2.11	5.74	1.87	.77	5.92	2.11	5.89	1.81	.81	.90
I	4.45	1.76	4.58	1.64	.20	4.77	1.88	4.77	1.62	.35	4.61	1.82	4.68	1.63	.28	.43
L	3.39	1.68	4.74	1.44	.18	3.23	1.75	4.29	1.49	.35	3.31	1.72	4.52	1.46	.26	.41
M	3.58	1.68	4.65	1.58	.41	3.13	1.41	4.71	1.73	.16	3.35	1.54	4.68	1.65	.29	.45
N	5.55	2.01	6.06	1.39	.24	5.61	1.96	5.55	1.88	.12	5.58	1.92	5.81	1.64	.18	.33
O	4.23	1.70	4.10	2.07	.66	3.71	1.61	3.58	1.64	.66	3.97	1.65	3.84	1.86	.66	.79
Q_1	4.74	2.00	5.06	1.46	.53	4.32	1.71	4.61	1.68	.49	4.53	1.86	4.84	1.57	.51	.68
Q_2	5.52	1.52	5.68	1.91	.75	5.58	1.85	6.23	1.95	.48	5.55	1.68	5.95	1.93	.64	.78
Q_3	6.61	1.47	6.84	1.87	.38	6.97	1.69	6.77	1.74	.38	6.79	1.78	6.81	1.80	.34	.51
Q_4	4.13	1.52	3.90	2.23	.52	3.90	1.91	3.26	1.61	.52	4.02	1.71	3.58	1.92	.54	.70

*For significance at the 1% and 5% levels of confidence, Pearson r's must equal .47. and .37.

**For significance at the 1% and 5% levels of confidence, Pearson r's must equal .34. and .26.

To avoid contamination effects the total sample was not pooled; instead correlations were computed between the 30 persons who took each of the two forms first and then for the 30 who took each of the two forms last. By means of Fischer's Z transformation [6], the two correlations were averaged.

***Computed with the Spearman-Brown Prophecy Formula [7]. These correlations are estimates of the reliability coefficients obtained by doubling the length of the test.

should ever be used alone, and further, that even when both tests are administered together, as recommended by the authors in their manual, the reliability of factors B, C, E, I, L, M, N and Q_3 is still low. In this connection it should be noted that in Supplement A to the test manual [4], the test authors themselves report that 10 of the 16 coefficients of equivalence for the 16 factors are less than .50, even after having been corrected by the Spearman-Brown formula to full length.

Having used the 16 P. F. Test as a part of the diagnostic psychological test battery in a clinical situation for over a year, it had been the authors' contention that factor B (general intelligence) was actually not a valid measure of intelligence, at least with the intellectually superior sample of patients seen in the clinic. This hypothesis was capable of being tested since scores from both the Wechsler-Bellevue Intelligence Scale, Form I (W-B I) and Wechsler Adult Intelligence Scale (WAIS) were available on 50 of the 71 patients who had taken the 16 P. F. Test. The results presented in Table 3 show that factor B, on both Form A and B, does not correlate significantly with the averaged Wechsler scores. The Verbal IQ's from the Wechsler tests were used in the computations in order to maximize the possibility of obtaining significant relationships. Evidently factor B is not a valid measure of intelligence in an intellectually superior sample.[3]

TABLE 3. CORRELATIONS BETWEEN FACTOR SCORES ON THE SIXTEEN P. F. TEST AND AVERAGED W-B I AND WAIS VERBAL I. Q. SCORES (N = 50)

Sixteen P.F. Test Form A Factors†	r* Verbal I.Q. with Form A Factors	Sixteen P.F. Test Form B Factors†	r* Verbal I.Q. with Form B Factors
A	.12	A	.01
B‡	.16	B‡	.19
C	.03	C	−.01
E	−.07	E	.05
F	.21	F	.14
G	−.26	G	−.05
H	−.15	H	−.05
I	.03	I	−.15
L	.12	L	−.16
M	.06	M	−.10
N	−.19	N	−.14
O	.18	O	.05
Q_1	.09	Q_1	.18
Q_2	.06	Q_2	.06
Q_3	−.07	Q_3	−.01
Q_4	−.05	Q_4	−.04

*Pearson r must equal .28 to be significant at the 5% level of confidence.

†For these 50 cases order of administration on the Sixteen P. F. Test was alternated in order to balance the effects of scale.

‡Factor B scores were also correlated separately for both forms with the following results: With Form A given first, the Pearson r between Verbal I.Q. with factor B on Form A was .22, while the correlation with factor B on Form B was .07. With Form B administered first, the Pearson r between Verbal I.Q. with factor B on Form A was .13 and the correlation with factor B on Form B was .30. (N for each group = 25).

As a further means of evaluating the construct validity of the 16 P. F. Test, each of the scores on the remaining 15 personality factors on Forms A and B was also correlated with the averaged W-B I and WAIS Verbal IQ scores. The results, shown in Table 3, demonstrate that none of the personality factors from either of the two forms

[3] When the 16 P. F. Test (Forms A and B combined) and the short form of the Otis Self Administering Tests of Intelligence, Higher Examination, Form C were administered to a random sample of 96 USAF basic airmen, the following coefficients of correlation were obtained: r Otis I.Q. with Factor A = −.16, B = .49, C = .28, E = .13, F = .24, G = .12, H = .06, I = −.33, L = .15, M = −.35, N = .18, O = .24, Q_1 = −.13, Q_2 = .13, Q_3 = −.16, and Q_4 = −.13. For significance at the 1% and 5% levels of confidence, Pearson r's must equal .26 and .20.

of the 16 P. F. Test correlates significantly with intelligence. These findings lend support to their construct validity as personality factors, and suggest that this test may be of use in certain selection situations in which nonintellective factors are importantly related to criterion measurements.

In order to determine whether experienced clinicians were capable of predicting which factors from the 16 P. F. Test would correlate significantly with each of the MMPI scales, five judges were used to make predictions. Two of the judges were psychiatrists and the other three were clinical psychologists. All of the judges had worked with the MMPI, while only 3 of the judges (the psychologists) had any experience at all with the 16 P. F. Test. Each of the judges was given a folder which contained an MMPI manual, the manual for the 16 P. F. Test, a sheet of paper on which each of the MMPI scales were listed, and a page of instructions which read as follows:

> "We are attempting to do a study which involves establishing the relationships between scales on the MMPI and those on the Sixteen Personality Factors Test having common psychological rationale. Your task involves taking the sheet of paper on which the fourteen (14) MMPI scales are listed and next to each of them putting the factors, or factor, from the 16 P. F. which you believe should be *significantly* related in either the positive or negative direction. Please indicate the direction you predict by placing a + or − after each factor. This should be done carefully, since on some of the scales of the 16 P. F. Test a high score does not have a negative connotation as it always does on the MMPI. Observe that on ten of the 16 factors (notably A, B, C, E, F, G, H, N, Q_2 and Q_3) high scores are desirable, while on the remaining 6 factors (namely, I, L, M, O, Q_1 and Q_3) low scores appear to be desirable. In every instance the direction you posit should be in terms of high scores on both the MMPI and 16 P. F. To facilitate this task you are being provided with a manual for each of the two tests."

The results presented in Table 4 demonstrate that none of the individual judges was able to predict a greater number of significant relationships than could be expected on a chance basis. Moreover, even when the selections were limited to those

TABLE 4. CHI SQUARE VALUES FOR EACH JUDGE

	Judges*				
	A	B	C	D	E
Number of correct predictions	13	8	12	15	20
Total number of predictions	29	13	20	33	55
Chi Square**	0.41	2.50	0.88	0.35	0.02

*The selections of each judge were compared with the actual number of significant correlations out of the total number of correlations computed between the two tests, i.e., 87 out of 224.

**With d.f. = 1, chi square must equal 3.84 for significance at the .05 level of confidence.

predictions which showed a minimum of 3 or more agreements among the 5 judges (10 correct out of 16 agreements), the Null hypothesis could not be rejected (X^2 = 1.60, P = >.20). These findings indicate the great need for further research on problems which give consideration to the clinician as well as to his measuring instruments [16, 17].

DISCUSSION

Special consideration should be given to certain relationships between the 16 P. F. Test and the MMPI which were expected to occur, but which did not. All five of the judges were in agreement in predicting significant correlations between D and factor F (depressive anxiety), Pt and factor C (ego strength), Ma and factor F, Pa and factor L (paranoid schizothymia), and Sc and factor A (cyclothymia). Only the

first three of these relationships were realized in the matrix of intercorrelations given in Table 1 while the last two were not. Four of the five judges predicted significant relationships between D and factor O (anxious insecurity), Pd and factors G (superego strength) and M (hysteric unconcern), and Mf and factor I (emotional sensitivity). All of the correlations predicted were realized except for that between Pd and G. Finally, three of the five judges concurred in predicting significant relationships between Hs and factor Q_4 (nervous tension), Hy and factors M (hysteric unconcern) and C (ego strength), Sc and factor H (adventurous resilience), Ma and factor A (cyclothymia), and Si and factors A and H. Of these seven predictions, four were correct, while the three between Hy and M, Hy and C, and Ma and A were not actualized. As Cronbach and Meehl [5, p. 300] have indicated, when "a predicted relation fails to occur, the fault may lie in the proposed interpretation of the test or in the network." Whatever the case, these findings indicate that some alteration in the "nomological network" of certain MMPI scales (Pa, Sc, Pd, Hy and Ma) and/or particular 16 P. F. factors (L, A, G, M and C) is necessary. It can, however, be stated that the present manner in which these scales and factors are described in their respective manuals does not enable experienced clinicians to make a significantly greater number of correct predictions about the relationships which exist among these measures than can be expected on a chance basis.

Some mention should be made of the possibility of developing measures of subjects' test-taking attitudes on the 16 P. F. Test. The findings presented in Table 1 suggest that there are enough factors on the 16 P. F. Test which correlate significantly with the L, F and K scales of the MMPI to make such an undertaking a distinct possibility. This is regarded as an important problem in clinical work since it is precisely with those patients who are reluctant to provide information about themselves that the busy clinician frequently turns to the personality tests for help. That the test authors are cognizant of some of the shortcomings of personality questionnaires when administered to subjects whose honesty, insight and cooperation are poor can be inferred from their statement that "results differ as between cooperative and uncooperative, well educated and poorly educated, honest and ulteriorly-motivated subjects." [3, p. 2]

Summary and Conclusions

Some aspects of the construct validity of the Sixteen Personality Factors Test were investigated by means of correlating its factors with the MMPI scales and with averaged Wechsler-Bellevue Intelligence Scale, Form I, and Wechsler Adult Intelligence Scale Verbal I.Q.'s. Coefficients of equivalence for each of the factors on the 16 P. F. Test were also reported. In addition experienced clinicians were asked to make predictions about which factors from the Sixteen P. F. Test would be significantly correlated with each of the scales of the MMPI.

It is concluded that (a) both the 16 P. F. Test and the MMPI measure certain common areas of personality; (b) the 16 P. F. Test has certain advantages over the MMPI, inasmuch as it contains fewer items, but apparently possesses greater breadth; (c) the coefficients of equivalence of the factors on form A and B are not high enough to warrant their being used as equivalent forms; (d) even when the coefficients of equivalence are corrected for full length by means of the Spearman-Brown prophecy formula, the resultant coefficients do not appear high enough to warrant the use of this test for accurate individual clinical diagnosis; (e) factor B is not a valid measure of intelligence in an intellectually superior sample of military flyers and appears to require redefinition; (f) some support for the construct validity of the other 15 factors (not including B) as measures of personality was obtained from the fact that none of these factors were significantly correlated with individual measures of intelligence; (g) certain scales on the MMPI (Pa, Sc, Pd, Hy and Ma) and/or particular factors on the 16 P. F. (L, A, G, M and C) appear to require redefinition, since in their present form experienced psychologists and psychiatrists are misled into making incorrect predictions about their interrelationships.

References

1. Cattell, R. B. *The description and measurement of personality.* New York: World Book Company, 1946.
2. Cattell, R. B. *A guide to mental testing.* Champaign, Illinois; Inst. Pers. and Abil. Test, 1602 Coronado Drive, 2nd Edition, 1948.
3. Cattell, R. B., Saunders, D. R., and Stice, G. F. *The Sixteen Personality Factor Questionnaire.* Champaign, Illinois: I. P. A. T., 1602 Coronado Drive, 1949.
4. Cattell, R. B., Saunders, D. R., and Stice, G. F. *The Sixteen Personality Factor Questionnaire, Supplements A, B and C.* Champaign, Illinois: I. P. A. T., 1602 Coronado Drive.
5. Cronbach, L. J., and Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
6. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1951.
7. Garrett, H. E. *Statistics in psychology and education.* Third Edition. New York: Longmans, Green, 1947.
8. Gilliland, A. R., and Colgin, R. Norms, reliability, and forms of the MMPI. *J. consult. Psychol.* 1951, 15, 435-438.
9. Hathaway, S. R., and McKinley, M. D. A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *J. Psychol.* 1940, 10, 249-254.
10. Hathaway, S. R., and McKinley, M. D. A Multiphasic Personality Schedule (Minnesota): II. A differential study of hypochondriasis. *J. Psychol.*, 1940, 10, 255-268.
11. Hathaway, S. R., and McKinley, M. D. A Multiphasic Personality Schedule (Minnesota): III. The measurement of symptomatic depression. *J. Psychol.* 1942, 14, 73-84.
12. Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* Revised Edition. New York: Psychological Corporation, 1943.
13. Hathaway, S. R., and McKinley, J. C. *Minnesota Multiphasic Personality Inventory manual,* Revised 1951. New York: Psychological Corporation, 1951.
14. Hathaway, S. R., and Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis: Univ. of Minnesota Press, 1954.
15. McKinley, J. C., and Hathaway, S. R. A Multiphasic Personality Schedule (Minnesota): IV. Psychasthenia. *J. appl. Psychol.*, 1942, 26, 614-624.
16. Karson, S., and Freud, S. L. Predicting psychiatric diagnoses with the MMPI. *J. clin. Psychol.*, 1956, 12, 376-379.
17. Meehl, P. E. Profile analysis of the Minnesota Multiphasic Personality Inventory in differential diagnosis. *J. appl. Psychol.*, 1946, 30, 517-524.

CONFIRMATION OF NINE PSYCHOTIC SYMPTOM PATTERNS

MAURICE LORR — *Veterans Administration and Catholic University*

JAMES P. O'CONNOR, JOHN W. STAFFORD — *Catholic University*

Introduction

The analysis to be presented is prompted by the belief that identification of the basic symptom patterns of the functional disorders represents an important and possibly necessary initial phase of research in this area. Once the basic dimensions of illness are isolated and quantitatively appraised, more fundamental research into the origin and dynamics of pathological and deviant behavior can proceed with greater certainty.

Dissatisfaction with current classificatory schemes is widespread. Initially achieved through acute clinical observation, a form of armchair factor analysis, diagnostic classifications are now periodically modified and "validated" by majority vote of practicing members of professional associations. This unsatisfactory state is in part due to the absence of specific known pathogenic organisms, characteristic organic brain lesions, or particular biochemical processes to account for mental illness. Vague definitions without upper and lower bounds are also partly responsible. However, much of the disagreement between experienced judges and the dissatisfaction in practice is accountable to the very considerable overlap of symptoms from one diagnostic category to the next[2]. As a consequence a large number of mixed types are encountered in practice.

In the domain of the functional psychoses five major studies have thus far been reported. Using Moore's[6] data, Degan[1] isolated some nine correlated first order and four second order factors. Wittenborn and his collaborators[8, 10] have repeatedly isolated some seven independent patterns. Lorr, Jenkins, and O'Connor[4] identified eleven first order and three second order factors. Recently Trouton and Maxwell[7] isolated six factors in a hospital Item Sheet. Guertin[3] from an unrotated cluster analysis of a check sheet identified five correlated factors for a sample of schizophrenics. Clearly there is need for cross-identification of the factors isolated and for repetition of such studies to determine the degree of invariance of the parameters postulated. Different sets of rating scales and check lists have been used to quantify behavior descriptions and different patient populations have been studied. Wittenborn *et al* and Trouton and Maxwell used orthogonal rotation; others have rotated to oblique simple structure. As a consequence there is question as to the number of unique factors and their relationships.

Accordingly in this study we sought to verify by means of multiple factor analysis 13 symptom and behavior dimensions previously identified as characteristic of functional psychotics. The hypothesized factors and the scale code numbers of the defining variables were as follows: (A) *retarded depression vs. manic excitement:* 1, 6, 11, 32, 41, 43, 50, 53; (B) *resistiveness:* 31, 35, 36, 46; (C) *paranoid projection:* 3, 13, 19, 42, 45, 52; (D) *activity level:* 28, 33, 37; (E) *melancholic agitation:* 2, 7, 15, 40, 48; (F) *perceptual distortion:* 8, 16, 24; (G) *motor disturbances:* 9, 12, 17; (H) *hostile belligerence:* 18, 29, 34; (I) *withdrawal vs. sociability:* 30, 38, 39, 54; (J) *grandiose expansiveness:* 14, 20, 25, 44, 49; (K) *conceptual disorganization:* 5, 10, 21, 51; (L) *obsessive-phobic reaction:* 22, 26, 47, 55; (M) *hysteric conversion:* 4, 23, 27.

Procedure

The total sample consisted of 116 functional psychotics from eight psychiatric hospitals. The sample from each hospital included open and closed ward cases, chronic and acutely disturbed cases, newly admitted and continued treatment cases. Alcoholics and neurological cases were excluded.

Ratings were made on a schedule of 55 scales, 39 of which were taken from the Multidimensional Scale for Rating Psychiatric Patients and 16 from the Wittenborn Psychiatric Rating Scales. These are the two best known standardized rating schedules. It seemed worthwhile to include representative scales from each form in order to determine their equivalence. Of the 39 MSRPP scales, 27 were taken from the interview section and 12 from the ward observation section. Variables numbered 1 to 39 are MSRPP scales; those coded 40 through 55 are Wittenborn scales. Both the MSRPP and Wittenborn scales were preceded by the full set of instructions provided by the authors.

Each rater completed the interview sections following a relatively thorough interview with the patient. The ward behavior section was completed either by an experienced attendant or nurse who knew the patient well, or by the interviewer on the basis of the report of the attendant or nurse.

Product moment correlations were computed between 52 of the scale variables. Three variables were dropped because a sizable number of patients were judged unratable. Since two of the variables omitted were introduced to define the hysteric conversion factor, the hysteric cluster could not be included in the study. When the correlations were segregated into 12 clusters, the cluster labelled hostile aggressiveness (H) was found to merge with the manic excitement cluster (A). The manic cluster and the 10 remaining clusters of correlations formed the basis for extracting 11 factors by the multiple group method. The resulting oblique matrix was orthogonalized and rotated by the single plane method to approximate simple structure. One rotated factor is considered residual and will not be described. Except where the pattern of correlations strongly suggests a lower limit, a cutting point of $\pm .35$ will be used in interpreting the factors.

Results

The scales correlated ±.35 or higher with Factor A are as follows:

No.	Scales			Correlation
29	swears less than others	—	swears more than others	.74
50	failure of affective responses	—	affective responses appropriate	.74
41	no evidence of loudness of voice	—	raises voice	.72
34	rarely irritable	—	more irritable than most	.70
53	responses slowed or delayed	—	no slowing of responses	.66
1	speech almost inaudible	—	speech distinctly loud	.65
11	conspicuously restrained in feeling	—	conspicuously unrestrained in feeling	.63
51	rate of speech uniform	—	rate of speech variable	.54
6	taciturn or mute	—	conspicuously over-talkative	.53
15	deeply depressed	—	highly elated	.47
2	placid or relaxed	—	distinctly tense	.47
54	neat and orderly	—	slovenly and unkempt	.38
18	bears no hostility	—	highly hostile	.36

Factor A represents a bipolar parameter; a retarded depression is at one pole and a manic excitement is at the other. A is very similar to Wittenborn's manic-depressed cluster [9]. It is broader than the original Factor A [4] since it includes the irritability, obscenity, and hostility of the hostile belligerence cluster. One pole seems to be identical with Degan's second order mania factor W which is defined by hyper-irritability and manic hyper-excitability [1]. It appears that Factor A is a second order factor appearing at the first order level.

The scales correlated ±.35 or higher with Factor B are as follows:

No.	Scales			Correlation
35	does expected without urging	—	requires urging to do what is expected	.46
46	reasonably cooperative	—	often reluctant to cooperate	.40
31	never resists expected activities	—	often resists expected activities	.36

Factor B is identical with the resistiveness factor previously identified [4]. Degan's neurasthenia factor I, defined by neurasthenia (appearance of fatigue) and negativism, bears considerable resemblance to B. The underlying element appears to be a stubborn reluctance to conform to or comply with the directions of others which is particularly characteristic of schizophrenic patients. A resistance to any form of energy expenditure may explain the presence of this factor in depressive patients.

The scales correlated ± .30 or higher with Factor C are as follows:

No.	Scales			Correlation
19	no unjustified suspicions	—	believes people talk about him	.65
3	no unjustified suspicions	—	believes he is conspired against	.58
52	does not feel misused	—	feels conspired against	.56
18	bears no hostility	—	highly hostile	.51
42	no evidence he feels controlled	—	believes others influence his behavior	.47
45	thinking not bizarre or obscure	—	thinking bizarre or obscure	.32
34	rarely irritable	—	more irritable than most	.32
16	no hostile impulses	—	disturbed by hostile impulses	.31

Factor C represents a paranoid type of projection clearly identical with Factor C of the original study [4], with Wittenborn's paranoid schizophrenia [9], and with Guertin's persecutory suspicious factor [3]. However, the hostility and the disturbance by impulses to commit hostile acts were not present in the original C. Degan failed to find such a factor in Moore's data for the simple reason that no variables defining the paranoid syndrome were explicitly included.

The scales correlated ±.30 or higher with and defining Factor E are as follows:

No.	Scales			Correlation
48	no complaint of anxiety	—	distressed by anxiety	.71
16	no hostile impulses	—	disturbed by hostile impulses	.49
22	no morbid fears	—	disrupted by morbid fears	.48
47	no obsessional thinking	—	obsessional thinking	.50
7	unconcerned	—	distinctly anxious	.46
20	no feelings of sinfulness	—	conviction of sinfulness	.43
40	no complaint of foreboding	—	feelings of impending evil	.35

4	no use of complaints	—	frequent use of complaints	.31
33	seems completely worn out	—	livelier than others	−.32
52	does not feel misused	—	feels conspired against	−.31
12	no manneristic movements	—	frequent manneristic movements	−.31
15	deeply depressed	—	highly elated	−.30
9	knows who he is and others	—	knows neither self nor others	−.36
10	thoughts and feelings consistent	—	thoughts and feelings unrelated	−.40

The parameter underlying E is a melancholy agitation. It is identical with Factor E of the original study [4], with the addition of morbid fears and obsessions. Wittenborn's anxiety factor is probably the same as E. Factor E is very similar to Degan's depression factor which is defined by depression, anxiety, refusal of food, tearfulness, and retardation of speech action. Only the retarded element in Degan's depression factor is inconsistent with the hypothesis of an anxious depression. The negative correlations of thought-feeling disharmony and manneristic movements with E were also present in the original E. These suggest that Factor E is actually bipolar, the negative end defining a disassociated lack of concern or apathy.

The scales correlated ±.35 or higher with Factor F are as follows:

No.	Scales			Correlation
4	reports no visions	—	sees and believes in visions	.66
8	reports no voices	—	hears and believes in voices	.62
9	knows who he is and others	—	knows neither self nor others	.46

Factor F was previously described as representing a pathological tendency towards perceptual distortion of reality. It seems identical with Degan's hyper-projection factor. Three conditions may have prevented its appearance in Wittenborn's data: orthogonal rotations, an insufficient number of reference variables, and incomplete reduction of the correlation matrix.

The scales correlated ±.35 or higher with Factor G are as follows:

No.	Scales			Correlation
12	no peculiar movements	—	frequent manneristic movements	.53
17	no bizarre postures	—	persistent bizarre postures	.49
54	neat and orderly	—	slovenly and unkempt	.37
7	unconcerned	—	distinctly anxious	.37

Factor G is essentially the same as the motor disturbance syndrome G previously isolated. Except for the absence of mutism and refusal of food, Factor G is the same as Degan's catatonia. In as much as few mute, stuporous patients were included in the present study it is not surprising that these elements are not represented on G. Wittenborn did not include scales defining manneristic postures or movements in his analyses. Motor disturbances, as exhibited in bizarre and repeated movements and postures, have been produced experimentally in animals and in humans by means of drugs. It is difficult to account for such disturbances at present, except as indicators of personality disorganization.

The scales correlated ±.35 or more with Factor I are as follows:

No.	Scales			Correlation
30	interested in nothing	—	many interests	.66
33	appears worn out	—	livelier than most	.63
28	actions slower than average	—	actions faster than average	.59
37	underactive	—	restless	.55
32	no answer if spoken to	—	hard to stop if spoken to	.39
38	stays by himself	—	usually with others	.37

Factor I appears to be a combination of previously isolated factors of social withdrawal versus sociability (I) and activity level (D). All elements of the original withdrawal factor except neatness are represented here. Similarly, every element defining activity level except distractiveness is represented on the new Factor I. Guertin's retardation and withdrawal factor is virtually identical with Factor I. Our hypothesis is that Factor I is a bipolar parameter of social withdrawal *vs.* excessive sociability. Further study with a larger number of cases should aid in testing this hypothesis.

The scales correlated ±.30 or higher with Factor J are as follows:

No.		Sacles		Correlation
5	no exaggerated opinion of self	—	grandiose convictions	.46
14	self depreciatory	—	self important or conceited	.30
34	rarely irritable	—	more irritable than most	-.31

Factor J represents an underlying self depreciation *vs.* grandiose expansiveness identical with the previously isolated Factor J. It is poorly defined in the present study but its meaning seems clear. In Wittenborn et al, this factor is called paranoid condition and is characterized by busyness, optimism, grandiose notions, and stilted speech. The per cent of total variance this factor accounts for tends to be small in all studies thus far reported.

The scales correlated ±.35 or higher with Factor K are as follows:

No.		Scales		Correlation
5	speech responses relevant	—	speech responses irrelevant	.69
21	speech coherent	—	speech scattered and incoherent	.62
43	ideas change normally	—	rapid or abrupt change of ideas	.53

Factor K is indicative of thinking disorganization. It appears to be the same as the factor called conceptual disorganization in the earlier study [4]. It was hypothesized that thought and feeling disharmony would help define this factor; this does not occur. A very similar syndrome is called schizophrenic dissociation by Degan because of specific disturbances in the association of ideas and incongruity of affective responses. Guertin's personality disorganization factor is very similar to Factor K. In a recent yet unpublished study, a rerotation of Wittenborn and Holzberg data [5] by one of the authors also revealed a conceptual disorganization factor. It was defined as irrelevant and incoherent speech, memory faults, variation in rate of speech, unawareness of the feelings of others, and inability to make simple decisions.

The scales correlated ±.35 or higher with Factor L are as follows:

No.		Scales		Correlation
55	no compulsive acts	—	compulsive acts occur daily	.55
26	no compulsive acts	—	resists performance of compulsive acts	.43

On the basis of the Wittenborn-Holzberg findings, a phobic-compulsive syndrome was hypothesized. Since Factor L is defined only by two very similar scales it may be that it is simply a doublet. Compulsive acts are not commonly found in psychotics. In fact a phobic-compulsive syndrome is ordinarily regarded as characteristic of neurotics. If few or no severe neurotics were included in the present sample studied we would not expect a phobic-compulsive syndrome. Furthermore, it is not uncommon for raters to confuse compulsive acts with manneristic stereotyped movements. Perhaps this is why compulsive acts helped define a catatonic excitement in the original Wittenborn analysis [8]. It may be that similar conditions resulted in Factor L in the present study.

Second Order Analysis. The correlations between the primary vectors were clustered and two factors were extracted simultaneously. The orthogonalized matrix was then rotated. Second order Factor X, shown in Table 1, evidently represents a manic excitement *vs.* retarded withdrawal with motor disturbances. A related factor, CC, identified as agitated hyperactivity *vs.* apathetic confused withdrawal was isolated in an earlier study [4].

Factor Y is defined at one end by grandiose expansiveness, paranoid projection, resistiveness, and a lack of concern. Self depreciation, a conforming cooperativeness, and a depressive agitation characterize the opposite pole. A similar paranoid-depressive Factor Y was identified by Degan. Only the grandiose element is missing from Degan's second order factor. The underlying psychological mechanism seems to be extropunitiveness *vs.* intropunitiveness. The patient either blames others for his difficulties or blames himself.

Discussion

Of the 13 syndromes hypothesized, 9 appear to have been confirmed. Hysteric conversion did not appear because two of the three defining variables could not be included in the study. Instead of a phobic-compulsive syndrome, a compulsive doublet emerged from the rotations. Activity level combined with social withdrawal, while hostile belligerence combined with manic excitement. Since the latter has also been identified by Degan as hyper-irritability, and by Guertin as excitement-hostility, it seems likely that some restriction in the nature and size of the sample accounts for its disappearance in this analysis.

Table 1. Rotated Second Order Factor Matrix
(Decimal points are omitted.)

First Order Factors				Second Order Factors X	Second Order Factors Y
A.	Retarded depression	—	manic excitement	54	41
B.	Cooperativeness	—	resistiveness	−13	63
C.	No projection	—	paranoid projection	−21	61
E.	No agitation	—	melancholy agitation	00	−55
F.	No distortion	—	perceptual distortion	05	14
G.	No disturbances	—	motor disturbances	−52	−34
I.	Withdrawal	—	sociability	72	02
J.	Self-depreciation	—	grandiosity	02	85
K.	No disorganization	—	conceptual disorganization	08	−01
L.	No conversion	—	hysteric conversion	00	−20
X.	Residual			66	−29

Each of the factors confirmed has now appeared in at least three independent analyses. Different rating schedules or check lists were used, and substantially different patient samples were studied. This suggests that the underlying parameters hypothesized may be used with considerable confidence. They represent useful constructs for describing and studying psychiatric patients.

Summary

This study sought to verify by means of multiple factor analysis 13 symptom dimensions previously identified as characteristic of functional psychotics. A sample of 116 psychotics from eight hospitals were described on 55 rating scales taken from two well known rating schedules. The scales were intercorrelated and factored by the multiple group method. Of the 13 syndromes hypothesized, 9 were confirmed with only minor modifications. Two second order factors extracted were also found to resemble factors previously isolated.

References

1. Degan, J. W. Dimensions of functional psychosis. *Psychometric Monogr.*, 1952, No. 6, 1-41.
2. Freudenberg, R. K. and Robertson, J. P. S. Symptoms in relation to psychiatric diagnosis and treatment. *Arch. Neurol. Psychiat.*, 1956, *76*, 14-22.
3. Guertin, W. A factor analytic study of schizophrenic symptoms. *J. consult. Psychol.*, 1952, *16*, 308-213.
4. Lorr, M., Jenkins, R. L. and O'Connor, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *J. abnorm. soc. Psychol.*, 1955, *50*, 78-86.
5. Lorr, M. and O'Connor, J. P. The relation between neurosis and psychosis: a reanalysis. Submitted for publication.
6. Moore, T. V. The essential psychoses and their fundamental syndromes. *Stud. Psychol. Psychiat.*, 1933, *3*, 1-128.
7. Trouton, D. S. and Maxwell, A. E. The relation between neurosis and psychosis. *J. ment. Sci.*, 1956, *102*, 1-21.
8. Wittenborn, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, *15*, 290-302.
9. Wittenborn, J. R., Herz, M. I., Kurtz, K. H., Mandell, W. and Tatz, A. The effect of rater differences on symptom rating scale clusters. *J. consult. Psychol.*, 1952, *16*, 107-109.
10. Wittenborn, J. R. and Holzberg, J. D. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, *15*, 372-380.

[1]Copies of tables showing the rotated oblique factor matrix V, arbitrary orthogonal matrix F and the transformation matrix may be ordered from the American Documentation Institute. Order Document Number 5131 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for photoprints or 35 mm. microfilm. Make checks payable to Chief, Photoduplication Service, Library of Congress.

THE LOYOLA LANGUAGE STUDY*

VINCENT V. HERR

Loyola University, Chicago

PROBLEM

The Loyola Language Study (LLS) is a modified form of the Kent-Rosanoff Association Test. Although free and controlled association tests have long been regarded as being among the most important instruments for measuring personality, very little actual research evidence has accumulated concerning their reliability and validity. Rosanoff *et al* [8] published the largest amount of data on the Kent-Rosanoff Association Test including tables of frequencies of popular and rare responses. Unfortunately, associative responses are largely determined by local and cultural meanings so that Rosanoff's frequency tables are of questionable value and validity for contemporary American groups. Malamud [7] conducted validity studies on Maller's controlled association test, and concluded that both validity and reliability were satisfactory, even for individual diagnosis, when proper systems of weights are used. Goodenough [5] recommended such tests for personality studies, and Crown [1] used the Maller test for distinguishing neurotics from normals. Our purpose has been to bring the association test up-to-date for current usage.

The problem in the present investigation was to ascertain whether the communality of the associations of hospitalized patients differed significantly from that of normals on the LLS. Specifically, schizophrenic responses would be expected to differ from normal in such a controlled test. Earlier investigations concentrated upon the reaction time for free association, and on the singularity of the responses which the subjects gave. The present investigation concentrates upon the fact of deviation from communality of response, and attempts to quantify these deviations. An essential prerequisite was to devise a satisfactory system of scoring communality of response without which small differences might be obscured.

METHODS

Test. The Loyola Language Study (Fig. 1) consists of eighty words chosen at random from the Kent-Rosanoff list in order to shorten the testing time to 30 minutes. The subjects were instructed to pause and consider what *most people* would be likely to think when confronted with each of the eighty stimulus words. The assumption was that psychotics in general, and schizophrenics in particular, would not be as capable of complying with this type of instruction and of finding common responses as the normals would.

Subjects. Since the Kent-Rosanoff frequency lists for free association were based on samples from the eastern United States, it was decided that the Loyola Language Study should be standardized separately for different areas. The present investigation began with normal and abnormal samples drawn from the Boston metropolitan area, was repeated in the Chicago area, and is projected for other areas in the United States.

Because age and education might affect the quality of the responses, a stratified random sampling technique was used in establishing norms for the general population. For the Boston experiment, the census statistics for the Boston metropolitan

*This project was sponsored in part by a grant from the Department of Welfare of the State of Illinois.

The author wishes to express his gratitude to the staffs of the two hospitals for their cooperation, and to all who acted as subjects and assistants. Particular mention must be made of L. B. Snider, J. Arsenian and O. Johnson who pioneered in this work; Mr. & Mrs. R. Lindberg who did most of the double and triple scoring; and of C. I. Doyle, F. J. Kobler, Mr. & Mrs. A. Delvecchio, and Mr. & Mrs. R. Stanek who carried out the Chicago project. H. J. A. Rimoldi was indispensible in devising scoring systems. Messrs. G Smola and V. Stewart did the intelligence studies, and J. Trainor collected the data on reliability.

FIGURE 1. THE LOYOLA LANGUAGE STUDY. DISCRIMINATING WORDS FOR CHICAGO MALES (M) AND FEMALES (F)

soldier (F)	high (F)	working	sickness (M)	dream (F)
hungry	white	comfort	mountain (M, F)	trouble (F)
butterfly	command	soft (M)	stove (F)	religion (F)
long	sour	short	girl	street
head (M, F)	king (M)	beautiful	salt	health
anger	deep	cold (M)	man	ocean
afraid	sleep (F)	whiskey	cheese	bed (F)
fruit	black (M)	yellow	baby	child
dark (M)	hammer (F)	window (M, F)	moon (F)	tobacco (M, F)
red	table (M, F)	scissors (M)	spider	woman
loud (F)	thirsty (M, F)	foot (M, F)	bread (M, F)	cabbage
bath	quiet	doctor (M, F)	whistle (M, F)	citizen (F)
eating	hard	wish (M)	carpet	earth
joy (M)	blue	house (F)	needle (M)	lion
rough (F)	sweet (M)	justice	hand (M)	butter (M)
heavy	stomach (M, F)	river	thief (M)	music

area for 1950 were consulted; counts were made of the numbers of persons in the general population with education at each possible level from fifth grade to sixteen years of education, and at each age level from nineteen years to sixty. Proportional numbers were used for the age and education sub-groupings in this study. Similarly, census statistics for the Chicago metropolitan area for 1950 were consulted; proportional numbers of subjects were chosen for the cells in the Chicago study. Four hundred males and 400 females were sufficient for a twelve-celled distribution, with four equal divisions with respect to education and three equal divisions with respect to age. However, since the frequencies of males differed from those of the females in any one particular cell, it was necessary to proceed with the two sexes separately in each area. Only persons who had been reared in the United States were included in the normative samples.

The hospitalized psychotic patients to be investigated in each geographical area were matched, person by person, with normals from the same area with regard to both age and education, because the possible effects of age and education were not known at the start. Patients were tested individually whereas normals were tested in groups of twenty to thirty. The psychiatric patients in the Boston area were inmates of a state hospital who had been diagnosed as predominantly schizophrenic. Seventeen percent of them were manics and two percent involutionals. The patients from the Chicago mental hospital had all been diagnosed as schizophrenics. In Boston, 50 male and 50 female patients served as subjects; in Chicago 56 males and 53 females. None of the patients had undergone shock treatment within three months prior to testing.

SYSTEMS OF SCORING

Systems of Scoring. Carter and Sorensen [1, 9] have already studied the problem of scoring free and controlled association responses, and have concluded that insignificant differences result from different methods of scoring. Three systems of scoring were employed in the present investigation.

In establishing norms for the general population, Boston males, Boston females, Chicago males, and Chicago females were treated as four separate categories. A frequency count was made, for instance, of the responses of the 400 Boston males to each of the 80 stimulus words. Thus it was established that, for example, 169 of the 400 Boston males responded to the word "hammer" with the word "nail". Percentages were then determined for each response to each stimulus word, and distributions of percentages were reduced by taking the square root of each. This procedure aligned all the responses of the normative sample on a descending scale, according to the amount of agreement among the normative group in choosing a particular response. The resultant psychological continuum, from low to high, based upon agreement of choices among the normative subjects, is the same kind of con-

tinuum involved in previous studies on communality of association response. After the mean root frequency and the standard deviation for each stimulus-response distribution had been calculated, standard scores were computed, with an arbitrary mean set at 20 and a standard deviation at 10. This method was thought to be justified because all the distributions were of the same shape and were skewed in the same direction. It had the advantages of an equal-interval scale, and independence of means and variances.

The final goal of scoring a test of communality must be to secure a single score for each subject for all his responses, a total score which may be taken as an indicator of the success that he has on repeated trials in finding common responses. Consequently, after the procedure described above had been carried out, test booklets of the 400 individuals in each of the four categories of subjects, namely the Boston men and women, the Chicago men and women, were scored according to the standard score scales determined from the frequency counts.

Although the standard score method just described had its advantages, it was costly in time and effort. A simpler method of scoring was also employed. This simpler method utilized double-root percentage-frequency values, so that the resultant scale values gave to each response to any given stimulus a value absolutely determined by its raw frequency. Thus no account was taken of the relationship between any given frequency and the other frequencies in the same distribution; no account was taken of the varying top frequencies from stimulus to stimulus. The Chicago samples of 800 normals and 109 patients were rescored by this method. A difficulty arose from the use of the absolute root-frequency scales. The variances turned out to be proportional to the means. This would preclude the use of certain statistical procedures with the data.

A still simpler type of scoring was attempted with the same 800 Chicago normals and 109 Chicago patients. Starting with the response of highest frequency to any given stimulus, as many other responses (in descending frequency) were accumulated to the highest response, as would be necessary to include the responses of 50% of the population. Any subject who chose a response that fell within the upper 50% of the normative sample was credited with a score of 1 for that item; any who did not choose a word from among them received a score of zero. For example, two responses to the stimulus "house", namely "home" and "family" together netted 51% of the total female normative group. Choice of either of these responses would merit a score of 1, and choice of any other would merit zero. When this method of "median" scores was used in scoring the eighty items, some patients scored near zero, and since it was planned to shorten the test later, it was not deemed advisable to work with the median scores. Actually the correlation between the standard scores and the doubled-root scores, for the same 800 subjects, was .96 and the correlation between the standard scores and the median scores was .93. Any of the three systems of scoring yielded differences between patients and normals which were significant beyond the .001 level of confidence. Yet somewhat different cutting points between patients and normals resulted if one system was used rather than another. The discussion of results which follows is based upon the standard scores only.

Results

The test booklets of 800 Chicago normals and 800 Boston normals, of 100 Boston patients, and of 109 Chicago patients, were scored with the standard scores. Then 53 Chicago female patients were matched for age and education with 53 normals selected from the 400 females; 56 Chicago male patients were matched with 56 normal Chicago males; 50 Boston female patients with 50 Boston normal females; and 50 Boston male patients with 50 Boston normal males. The matching was person to person, and the limits of the differences allowed were one year of age and one year of education. The mean difference between normals and patients was in each case significant beyond the .001 level of confidence. But there was considerable overlap on the extremes. In order to ascertain which items contributed most to the

differences, an item analysis was made on all 80 items, comparing the mean score for each item made by the normals with that made by the patients. At least 25 items were significantly different at the .01 level of confidence for each of the four comparisons. The booklets were then rescored using only the 25 items which showed the maximum difference. After computing new means and standard deviations, it became apparent that the shortened test had a much higher screening efficiency than the longer test. Fig 1 contains the list of the 80 stimuli and the top 25 items for Chicago men are marked with M, those for the Chicago women, with F. Only eleven of the top items are common to both Chicago men and women. The same is true for the Boston men and women. Just half of the top screening stimuli are common to any one sex, regardless of the area of residence.

Validity Coefficients. An index of screening efficiency, according to a method devised by Eysenck [3, 4] was computed for the total scores on the shortened tests. It consists in subtracting the proportion of normals incorrectly identified from the proportion of patients correctly identified. The indices are as follows: Chicago men .64, Chicago women .70; Boston men .70, Boston women .71.

Another method of estimating validity was used. Phi coefficients for each of the 25 items were computed and converted into product-moment coefficients, on the assumption of continuity between normals and patients on communality of thought. The median coefficients were as follows: Chicago men .45, Chicago women .66; Boston men .65, Boston women .68. The shortened test has concurrent validity, since it distinguishes the patients from the normative groups.

As a final check on validity, the ratings of psychologists on 50 Boston female patients were secured from the hospital. Ratings ranged from 0 - 80 and were nearly normally distributed. The product-moment coefficient of correlation between the ratings and the scores on the shortened test was .48, which is significant beyond the .01 level of confidence. The shortened test seems suitable for differentiating between various degrees of mental illness.

Reliability Studies. Reliability was estimated in two ways. The internal consistency was measured by correlating the odd-numbered items with the even-numbered items for samples of 400 men and 400 women in groups of 100 each. The split-half reliability when corrected by the Spearman-Brown formula ranged from .88 to .94 with a median at .92.

The test-retest reliability was also considered. This was obtained by administering the test twice to the same persons, at intervals of time from 4 to 8 weeks. The correlation for eighty items was .49, and this value did not change when the odd items of the test were correlated with the even items of the retest. When the total length of the test was reduced by scoring only the items which had top screening efficiency, the correlation between test and retest rose to .55 for a sample of 82 male and female subjects. Other groups are being tested in order to confirm this reliability figure. Guilford [6] tells us that both reliability and validity cannot be maximum in a test which has items with a variety of difficulty and there must be a compromise between the two. The values for both presented in this study fall well within the suggested ranges for tests of this type.

The factors of age and education might be thought to influence the performance on the LLS. In order to offset the influence of these variables, matched groups were used for the item analysis of the test protocols. The four total stratified samples of 400 each were used and a coefficient of correlation computed between age and the Language score. The results are as follows: Chicago men, −.22, Chicago women, −.23; for 400 Boston men, −.29, and Boston women, −.31. The relationship between age and the scores was not rectilinear, and the coefficients varied from one subgroup to another. As an example, taking the group which had three years of high school, age correlated with the LLS −.31 for the Chicago men; but taking the group which had two years of college, the correlation was −.12. The coefficients are slightly lowered by using only the top 25 items in scoring. With these reduced scores, tables are

being prepared which will enable investigators to make the small corrections for age variance.

On the other hand, the LLS correlated positively with years of education and the coefficients are as follows: Chicago men .21, Chicago women .22; Boston men .17, Boston women .18, and all these coefficients are significant at the .01 level of confidence. However, the general trend observable among the normals for education to be related to better LLS scores is not found among the patients. For some unknown reason they neither show consistent losses due to age nor consistent gains due to education. The present samples were too small to give significance data.

The intelligence factor was also considered in the present investigation. No intelligence scores were available for the large stratified samples, but the positive correlation already mentioned between education and the LLS score suggests a relationship between intelligence and the LLS. In order to eliminate the possible effect of varying amounts of education on the LLS, two homogeneous groups were selected for study. The one consisted of 100 senior high-school boys who had been tested with the Henmon-Nelson test, Form C; the other of 100 male college freshmen and 100 female college freshmen who had been tested with the American Council on Education entrance examination. For the high-school senior boys, the correlation with the LLS was .11, significant at the .30 level, whereas for the college males it was .06 and for college females , .04. The correlation between the "L" score on the ACE and the LLS score was .06 for the males and for the females. Apparently intelligence apart from education is a negligible factor in communality of thought.

Summary and Conclusions

The Loyola Language Study distinguishes between schizophrenics and normals, using any of three systems of scoring, namely standard scores, doubled-root frequency scores, and median scores. The screening efficiency of the test is increased by using only high validity items scored by standard scores. Such scores also permit of parametric statistical operations. There is a tendency for older persons to make poorer scores, and for this effect to be counteracted by education. Intelligence, apart from age and education, is not a significant factor.

References

1. Carter, H. D. A preliminary study of free association: I. Twin similarities and the technique of measurement. *J. Psychol.*, 1938, *5*, 201-215.
2. Crown, Sidney. The word connection list as a diagnostic test; norms and validation. *Brit. J. Psychol.*, General Section, 1952, *43*, 103-112.
3. Eysenck, H. J. A comparative study of four screening tests for neurotics. *Psychol. Bull.* 1945, *42*, 659-663.
4. Eysenck, H. J. Screening out the neurotic. *Lancet*, Apr. 19, 1947.
5. Goodenough, Florence. The use of free association in the objective measurement of personality. Chapter V. in McNemar, Q., and Merrill, M. A., Eds. *Studies in Personality*. New York: McGraw-Hill, 1942.
6. Guilford, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1950, p. 523.
7. Malamud, D. I. Value of the Maller controlled association test as a screening device. *J. gen. Psychol.*, 1946, *21*, 37-43.
8. Rosanoff, A. J. (ed). *Manual of Psychiatry*, Sixth Edition. New York: Wiley, 1927.
9. Sorensen, M. I., and Carter, H. D. Twin resemblance in community of free association responses. *J. Psychol.*, 1940, *9*, 237-246.

MEASURING THE IMPAIRMENT OF THE ABSTRACTING FUNCTION WITH THE PROVERBS TEST[1]

CLYDE M. ELMORE AND DONALD R. GORHAM

Headquarters Flying Training Air Force, Waco, Texas

Veterans Administration Center and Baylor University, Waco, Texas

Problem

Many writers have called attention to the problem of language disturbance associated with psychiatric disorders[1, 2, 6, 11, 12, 13, 16]. Goldstein and others[1, 3, 4, 7, 15] have shown that impairment of the abstracting function is found in both schizophrenia and organic involvement. Psychiatrists have long associated proverbs with the measurement of verbal abstraction, leaning heavily upon Benjamin's[3] list.

In 1956, Gorham[8, 9] reported a Proverbs Test which offered three parallel forms of a clinical test with an objective method of scoring including norms for normals and schizophrenics. A multiple-choice form of his test made group administration and clerical scoring possible. Gorham[10] demonstrated that either form of the test would differentiate a normal and chronic schizophrenic population of male veterans at a high level of statistical significance. This finding raises two questions: Are the results equally applicable to male and female populations? Would this apparent indication of impairment in abstract thinking be reflected by a greater loss if the test were applied to organic cases? It is the purpose of this study to use the Proverbs Test to test the hypothesis that there is a drop in level of the abstracting function from normal to chronic schizophrenic to organic subjects. Both sexes will be investigated.

A Description of the Proverbs Test

The clinical Proverbs Test consists of three parallel forms containing 12 proverbs each, scaled in order of difficulty. Samples are:

1. Where there's a will there's a way.
2. Barking dogs seldom bite.
3. A stream cannot rise higher than its source.
4. Let sleeping dogs lie.
5. The wife is the key to the house.
6. Great bodies move slowly.

An abstract (Ab) score on this form of the test is secured by rating answers according to sample scores provided for each proverb. Each answer is rated 0, 1 or 2 points; the total possible score being 24.

The multiple-choice Proverbs Test consists of forty items with four answers each. It is scored by stencil. An abstract (Ab) score consists of the total number of correct responses. A concrete (Co) score consists of the number of items marked from 18 "concrete" answers. Samples of an easy and a difficult item are:

1. The sun shines on all alike.
   a. It's the same sun everywhere.
   b. All are created equal. (Ab answer)
   c. The sun shines on everybody. (Co answer)
   d. People that do the same things are alike.
2. Crooked logs make a straight fire.
   a. Even the humblest can create things of beauty.
   b. Handicaps may be used to advantage. (Ab answer)
   c. There is some good in everybody.
   d. Things will come out all right in the end.

The clinical manual[9] for the test provides norms for schizophrenic and normal populations including mean scores by intelligence levels.

---

[1]This research was supported by a grant from the Hogg Foundation for Mental Hygiene, Austin, Texas. The study was submitted by Mr. Elmore in partial fulfillment of the requirements of the M. A. degree at Baylor University.

Design

The essential design of this study consists of matching cases for intelligence in order to compare Proverbs Test scores by sex and diagnostic group. The index of intelligence used was the Word Knowledge Test of the Airman Classification Battery.[2] Individual male and female subjects of chronic schizophrenic and organic diagnoses were matched within one stanine score on the vocabulary test with normal subjects. The patient population was secured from Texas state hospitals. Those selected had been hospitalized for more than five years and had not received active treatment (electro-convulsive, insulin coma, or tranquilizing drugs) within the prior year. The organic group had been diagnosed as "organic without functional psychosis" and otherwise met the same specifications as the schizophrenic group. The control population was drawn from a large normal population of industrial and office workers, laborers, etc.

Results

Three measures of abstraction were secured from the two forms of the Proverbs Test; an abstract score and a concrete score from the multiple-choice test and an abstract score from the clinical test. The mean scores on each of these measures is shown in Table 1. This table confirms the hypothesis of this study: that the ab-

Table 1. Mean Scores on the Proverbs Test (N = 25 for All Groups)

Population Group	Multiple-choice test scores		Clinical test Form I
	Abstract score	Concrete score	Abstract score
Normal			
Males	22.7	1.8	11.4
Females	22.5	1.6	9.4
Schizophrenic			
Males	12.0	4.9	1.6
Females	14.2	3.3	2.5
Organic			
Males	11.3	6.8	1.2
Females	12.5	6.5	1.9

stracting function drops from normal to schizophrenic to organic subjects. It will be noted that the concrete scores increase as the abstract scores decrease. It is assumed by the authors that these two measures are representative of two aspects of the same intellectual process, viz., verbal abstraction. "Concrete" thinking is considered to be the opposite of, or the inability to do abstract thinking. This assumption is substantiated by the negative correlation (−.58 for 100 normal subjects and −.54 for 173 chronic schizophrenic subjects) found by the author in a previous study (10). Further statistical analysis of the data indicates which scores differentiate various diagnostic groups with statistical significance.

Multiple-Choice Abstract Score. The variation in the abstract scores on the multiple-choice test was analyzed by sex, vocabulary and diagnostic grouping. The findings are presented in Table 2 and show no significant sex differences, that is patients or normal subjects of either sex make approximately the same scores on the Proverbs Test whether divided by vocabulary level or diagnostic category. It can also be inferred that the diagnostic categories are independent of intelligence as measured by vocabulary level. The close relationship of vocabulary level and the Proverbs abstract score confirms Gorham's (8) earlier findings based solely on male military personnel. The important finding in terms of this investigation is the highly significant variation of Proverbs abstract scores between diagnostic categories which are independent of sex and vocabulary level. This finding directly supports the hypothesis

[2]Used by permission of the AF Personnel and Training Research Center, Lackland Air Force Base.

TABLE 2. ANALYSIS OF VARIANCE FOR ABSTRACT SCORE ON THE MULTIPLE-CHOICE PROVERBS TEST

Source of Variation	d.f.	Mean Square	F
Sex	1	54	1.9
Diagnostic category	2	1651	59.2*
Vocabulary level	24	206	7.4*
Sex x Diagnosis	2	14	0.5
Sex x Level	24	37	1.3
Diagnosis x Level	48	29	1.0
Sex x Diagnosis x Level	48	28	
Total	149		

*Significant at the .001 level of confidence.

that schizophrenia and organic involvement result in impairment in the abstracting function.

To determine where the variation existed between the three diagnostic categories, the *t* test was applied with values as follows: normals and schizophrenics, 7.21; normals and organics, 6.53; schizophrenics and organics, .47. The first two *t* values are significant at the .001 level of confidence while the third one is obviously not significant. These data indicate that the abstract score of the multiple-choice test tends to show a striking impairment of abstraction between normals and both psychiatric groups. It does not supply any evidence of a greater loss for the organic group.

Multiple-Choice Concrete Scores. The application of Friedman's chi-square test to the concrete scores yielded a chi-square value of 14.23 which is significant at the .001 level of confidence. The *t* test was considered to determine which groups could be separated by the concrete score. Because of the large number of zeros this test could not be applied to the normal group. Fortunately, however, the *t* test could be used in comparing the scores of the schizophrenic and organic groups. The value of *t* was 5.2 which is significant at the .001 level of confidence. The concrete score, therefore, supplies the needed index for differentiating these two groups which could not be distinguished by the abstract score.

The Clinical Test: Form I. Analysis of the scores of the clinical test was made difficult by the high incidence of zero scores among both schizophrenic and organic patients. The test scores were ranked and the Friedman chi-square test was applied with a resulting chi-square value of 43.1, which is significant at the .001 level of confidence. An inspection of the mean scores of the clinical test in Table 1 is sufficient to reveal that the differentiation in mean scores is between normals and organics or schizophrenics. This test might have some value, then, in distinguishing normals from chronic psychiatric patients. It throws little or no light on the hypothesis being investigated in this study.

DISCUSSION

It was the assumption of this study that the clinically discernable impairment of the abstracting function found in schizophrenic and organic patients could be demonstrated quantitatively by a test of verbal abstraction. The results of the study have justified this assumption. The experimental design provided a rigid test of the hypothesis, since the factors of sex and intelligence were controlled in order to isolate the variation which was due solely to diagnostic grouping.

The two scores, abstract and concrete, of the multiple-choice test are most useful in differentiating between the three diagnostic groups. Either of these scores will differentiate at a high level of confidence between the normal and either of the two psychiatric groups. The concrete score makes the added contribution of separating the schizophrenic and organic groups at a high level of confidence. Since this form of the test is suitable for group administration and scoring by stencil, the potential usefulness for research of this measurement of the abstracting function is suggested.

A point of theoretical interest is the negative relationship between the abstract and concrete scores of the Proverbs Test. This fact strongly suggests that the abstracting function and concreteness of thinking may actually be opposite portions of an abstract-concrete continuum, rather than two dichotomous thinking qualities. The findings of this study point to this type of modification of intellectualization in the process of impairment which occurs in certain cases of schizophrenics and organics. The abstracting attitude of normals apparently gives way to a greatly diminished abstraction with associated concreteness in chronic schizophrenics compared with almost complete substitution of concreteness for abstraction in organics.

The clinical form of the test does not lend itself as well for this type of study, partly because a concrete score is not readily derived and partly because the material elicited by this test is most significant when considered qualitatively as well as quantitatively.

Conclusions

1. A rigid experimental design was used to test the hypothesis that clinically discernable impairment of the abstracting function could be measured quantitatively. Normal, schizophrenic and organic subjects were matched for intelligence and compared by sex and diagnostic groupings.

2. The abstract score of the multiple-choice Proverbs Test differentiated between normals and either psychiatric group at the .001 level of confidence. It failed to differentiate organics from schizophrenics.

3. The concrete score of the multiple-choice Proverbs Test differentiates between normal and psychiatric patients and in addition distinguishes between chronic schizophrenic and organic groups at the .001 level of confidence.

4. The findings of the study confirm the hypothesis that a drop in level of the abstracting function from normal to chronic schizophrenic to organic groups can be demonstrated quantitatively.

References

1. Arieti, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.
2. Baumgarten, F. A proverb test for attitude measurement. *Pers. Psychol.*, 1952, 5, 249-261.
3. Benjamin, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univ. of Calif. Press, 1944. Pp. 65-90.
4. Bolles, M. and Goldstein, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, 12, 42-65.
5. Fruchter, B. The factorial content of the Airman Classification Battery. *Res. Bull.*, 1949, 49-1. San Antonio, Texas, Lackland Air Force Base.
6. Goldstein, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univ. of Calif. Press, 1944. Pp. 17-40.
7. Goldstein, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, 96, 575-88.
8. Gorham, D. R. A proverbs test for clinical and experimental use. *Psychol. Rep. Monogr. Suppl.* 1956, No. 1.
9. Gorham, D. R. Proverbs Test. Clinical Forms I, II and III. Best Answer Form. Manual. Clinical Manual. Louisville, Ky.: Psychological Test Specialists, 1956.
10. Gorham, D. R. Use of the Proverbs Test for differentiating schizophrenics from normals. *J. Consult. Psychol.*, 1956, 20, 435-440.
11. Hanfmann, E. and Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
12. Hayakawa, S. I. *Language in Action.* New York: Harcourt Brace, 1941.
13. King, W. R. Ability to abstract. In F. A. Mettler (Ed.), *Selective partial ablation of the frontal cortex: a correlative study of its effects on human psychotic subjects.* New York: Hoeber, 1949. Pp. 218-38.
14. Loeb, B. The function of proverbs in the intellectual development of primitive peoples. *Sci. Mon.*, 1942, 74, 100-104.
15. Meadow, A., Greenblatt, M. and Solomon, H. C. "Looseness of association" and impairment in abstraction in schizophrenia. *J. nerv. ment. Dis.*, 1953, 118,27-35.
16. Moran, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, No. 370.

EVALUATION OF INTELLECTUAL POTENTIAL IN HEMIPLEGIC INDIVIDUALS

FRANKLIN C. SHONTZ[1]

Highland View Hospital, Cleveland, Ohio

PROBLEM

Practically none of the currently popular psychological tests provides norms for the evaluation of hemiplegic individuals; and, if they did, such norms would probably be practically useless since almost all available tests of intellectual capacity utilize either verbal communication or performances requiring manual dexterity on the part of the subject. The construction of specialized norms, therefore, would often imply a contamination of the underlying factor being measured with the physical effects of the disability itself. For the psychologist this currently means it is practically impossible to make statements about a physically disabled person which clearly differentiate between the person's ability to express his capacities and the underlying extent of the capacities themselves as potential psychological resources. Further, it is currently impossible to make statements about the disabled individual in terms of his own group. Both are serious problems in a rehabilitation setting, because actual rehabilitation potential probably depends far more upon internal capacities than upon the physical ability of the patient to express himself intelligently on typical psychological tests. Success in a rehabilitation program, designed by experience to meet the needs of the physically disabled, probably does not require what would usually be called a "normal" level of abilities. How well a hemiplegic performs in rehabilitation, for example, may be much more closely determined by his abilities relative to other hemiplegics, than by his abilities relative to the general population.

The hemiplegic (especially the right hemiplegic) is in a particularly difficult situation. Organic brain damage affects both his intellectual functioning and his physical means of expression. The right hemiplegic has usually lost the functioning of his dominant body side and may, in addition, have particularly disturbing problems with verbal expression. Usually on superficial examination, right hemiplegics appear to be more seriously disturbed than do left hemiplegics. Therefore, it is of especial importance that our evaluation of these people be in terms of norms based on their own group and of tests which minimize their physical, though not their intellectual, loss.

METHOD

Procedure. Two tests usually used for children were administered to a sample of 50 mixed hemiplegic patients. These tests require only gross arm movements as responses, thereby minimizing the effect of the hemiplegic's loss of dexterity and possible verbal-expressive difficulties. One test, the Knox Cubes from the Arthur Point Scale of Performance tests (1), requires the subject to repeat tapping patterns demonstrated to him by the examiner on a set of 4 stationary blocks. No speech is used and the tapping itself requires only gross arm movements. Primarily, this test is thought to measure immediate memory, or recall. The other test, the Columbia Mental Maturity Scale (2), was designed for use with handicapped children. It requires the subject to pick from successive series of pictures the one out of three, four, or five stimuli which "does not belong" with the others. The task, therefore, seems to involve abstraction and discrimination, although it is often presented as a measure of intelligence.

The Knox Cubes series were administered twice, once preceding, and once following the Columbia Mental Maturity Scale. Both tests were scored in the manner prescribed by their originators.

[1]Thanks are due to Marilyn Weston and Robert Postel for their assistance in collecting data for this research. Additional data available from the author on request.

Subjects. The sample was composed of 50 institutionalized hemiplegic patients, representing almost all the testable and clearly hemiplegic patients in a 450 bed hospital for the chronically ill. Of the 50 subjects, 17 were right hemiplegics (10 males, 7 females), and 33 were left hemiplegics (17 males, 16 females). The mean age was 63.0 years with a standard deviation of 9.47. There was no statistical relationship between the side affected by the hemiplegia and either age or sex; nor was there any relationship between age and sex in the sample. The group was somewhat heavily loaded with left hemiplegics; but, whether this is representative of the general population of hemiplegics is not known.

The Knox Cubes and Columbia Mental Maturity Scale were administered to all 50 subjects. In addition, data were available for 31 of these subjects on four subtests of the Wechsler Memory Scale: Mental Control, Logical Memory, Digit Span, and Associate Learning. Five of these 31 subjects were so verbally disturbed, however, that they could not score over 2 points on all four Wechsler subtests. Therefore it was necessary to determine all relevant results both with and without these five subjects. In general, their inclusion exaggerated all other findings slightly, probably by virtue of their extreme position with respect to the rest of the distributions. There was no evidence that the sub-group taking the Wechsler Memory Scale was in any other way different from the sample as a whole.

Results

The Columbia Mental Maturity Scale and the Knox Cubes tests were found to correlate about .60 with each other, the correlation varying from .58 to .61, depending upon the group or sub-group involved. This correlation is significant at a probability level of .01.

The Columbia Mental Maturity Scale and Knox Cubes correlated .47 and .42, respectively with the subtests of the Wechsler Memory Scale. Both correlations are significant at the .05 level, and both correlations dropped slightly below the level of statistical significance when the 5 extremely disturbed cases were removed from the sample tested. No significant differences were found between right and left hemiplegics on the Columbia Mental Maturity Scale or the Knox Cubes.

Right hemiplegics were found to obtain significantly lower scores ($p < .03$) than left hemiplegics on the combined subscales of the Wechsler Memory Scale. Removing the Digit Span subtest, however, reduced biserial correlations to chance levels and eliminated the obtained difference between groups. No significant relationship was found between age and performance on any of the three research measurements.

Frequency distributions for the Columbia Mental Maturity Scale and the Knox Cubes tests were substantially normal, although the distribution for the Knox Cubes test skewed slightly to the left. In approximate mental age terms, both groups averaged about seven years, and the Knox Cubes gave a somewhat greater spread of scores than did the Columbia. These results fall about four and one-half years below what an extrapolation of Wechsler's approximations[3, p. 226] might lead one to expect for this age-group. Statistics describing these frequency distributions are presented in greater detail in tables 1 and 2.

Table 1. Percentile Distribution of Knox Cubes and Columbia Mental Maturity Scale Scores

Percentile	Raw Score (Knox Cubes)	Raw Score (Columbia Mental Maturity Scale)
90	11.25	69.5
75	10.25	63.5
50	8.25	55.5
25	5.50	49.0
10	3.25	44.0

TABLE 2. DISTRIBUTION OF Z-SCORES FOR KNOX CUBES AND COLUMBIA MENTAL MATURITY SCALE

Z-Score	Raw Score (Knox Cubes)	Raw Score (Columbia Mental Maturity Scale)
+2.0	13.25	77.0
+1.0	10.50	66.5
0.0	7.75	56.0
−1.0	5.00	45.5
−2.0	2.25	35.0

CONCLUSIONS

The Columbia Mental Maturity Scale and the Knox Cube tests are appropriate instruments for the evaluation of the intellectual capacities of hemiplegic patients. They do not require the use of expressive language, nor do they make demands for a high level of manual skill. They do not seem to be affected by the side on which the hemiplegia occurs, and the materials were not found to be insulting to the adult intellect. None of the subjects in this sample were found to be untestable on these examinations, and both tests are easy and quick to administer. The tests are highly enough correlated to warrant the assumption that both measure the same thing, yet the correlation is not sufficiently high to allow for prediction in the individual case.

On the other hand, the Wechsler Memory Scale does seem to penalize certain subjects with verbal-expressive problems; and it tends to be more difficult for right than for left hemiplegics, especially the Digit Span sub-test. A few subjects were found to be almost completely untestable with the Wechsler subtests, yet these same subjects were found to demonstrate measurable intellectual and memory ability on the other two examinations.

REFERENCES

1. ARTHUR, G. *A point scale of performance tests.* New York: Psychological Corp., 1947.
2. BURGEMEISTER, B., BLUM, L. H. and LORGE, I. *Columbia mental maturity* scale. Yonkers-on-Hudson: World Book Co., 1954.
3. WECHSLER, D. *Measurement of adult intelligence.* Baltimore: Williams & Wilkins Co., 1952.

THE PREDICTION OF ADJUSTMENT OVER A FOUR YEAR INTERVAL*

HARRY BEILIN

Institute of Child Welfare, University of Minnesota

Introduction

This paper reports one aspect of the Nobles County (Minn.) Mental Health Survey which has been underway since 1949. The general investigation had three main aims: (1) to undertake a study of the prediction of adjustment over time, (2) secure age data on personality and adjustment measures, and (3) develop screening instruments for locating children and adolescents in need of special attention. The first of these aims, the prediction of adjustment, will be reported on here. The predictions cover a four year period and the results to be reported are for the men in the sample. The problems of the present prediction study, can be reduced to (1) the development of predictors, (2) the development of criteria, and (3) the relation of predictors to criteria.

Methods

Development of Predictors. In studies of the prediction of mental health, difficulties arise from the lack of clarity and diversity among conceptions of its nature and from the variety of approaches to its measurement. Variables effective as predictors may be quite different in character from the phenomena being predicted. This is particularly true if the predictor variables are not operative over the period from prediction to validation and if behavioral discontinuities exist to encourage what might be evaluated as desirable or undesirable at one period of time (*e.g.*, childhood or adolescence) to be judged the opposite in another period (e.g., adulthood). As an illustration, dependence upon parents is a desirable feature of childhood but undesirable if characteristic in adulthood.

With knowledge of "age-related" and "age-free" behavior quite limited, the choice of relevant predictor variables is facilitated in one of two ways or in a compromise of these: in the framework of a theory or by strictly empirical means. In the present study, the predictor instruments were designed to reflect, in general, the "personal" and "social" adjustment of the subject at the time of testing. The behavior areas included were based on previously established relationships or assumptions having a fair measure of face validity. The instruments can be classed in two groups: those to which the subject responds and those employed by his teachers to evaluate his behavior. The 1950 predictor instruments completed by the pupils were of the inventory, sentence completion, rating scale and nominating form types. They include measures of the child's relations to his family, social responsibility, personal habits and skills about the home, play and recreational activities and interests, and a scale of psychoneurotic items.

The instruments employed by teachers included a nomination form (of the best and poorest adjusted), a Thurstone type rating scale of responsible and irresponsible behaviors, and a personality rating scale.

In 1950 the psychological instruments developed or adapted by the Institute staff were administered to 3200 children and adolescents between the ages of 9 and 18 in all the schools (public and parochial) of a Minnesota county. Ratings by teachers were obtained from the administration of the aforementioned instruments.

---

*This study has been supported in part by research grant M-690 from the National Institute of Mental Health, U. S. Public Health Service, and in part by funds from the Institute of Child Welfare, University of Minnesota. The principal investigators are J. E. Anderson and D. B. Harris. The author's tenure with the study was from 1953-56.

A major portion of this report was presented in the symposium "The Prediction of Adjustment" at the 1956 meetings of the Midwestern Psychological Association in St. Louis.

Concurrent validation of the 1950 measures was accomplished by using as criteria the Juvenile Court judge's nominations of delinquents and teachers' descriptions of these youngsters. The instruments as a group were sufficiently discriminating, employing these criteria, to suggest the value of predictive validation. For prediction two indices were developed. The *Pupil Index* is a composite score derived from the group of instruments to which the youngster responds.[1] The *Teacher Index* is a composite score resulting from four teacher's rating instruments. In the prediction study, high standard scores (7.0 - 9.9) have been used to predict good adjustment, low scores (0.1 - 3.0) to predict poor adjustment, and middle scores (4.0 - 6.0) to predict adequate, but neither good nor poor adjustment. In this study poor adjustment is more broadly conceived than as involving neurotic or psychotic behavior. It includes, for example, behavior which might involve the individual with the law or with therapeutic or custodial agencies.

Development of Criteria. In selecting the conceptual criteria for the present study [4] we were interested in assessing the following: (1) absence of mental disease or behavior disorder, (2) normality of behavior (*i.e.*, social adjustment, adjustment to environment, *etc.*), (3) unity of personality (*i.e.*, acceptance of self, ego strength, value development, etc.), (4) adequate perception of reality (of self and external world), and (5) developmental resolution of conflict (the epigenetic principle of Erikson) [3]. Our view has been that no *one* operational criterion would be adequate to reflect these essentially conceptual criteria and as a result a variety of criteria have been employed in the investigation as described below.

1. *The interview criterion.* The interview samples a number of areas germane to individual adjustment including: school adjustment, use of leisure, job adjustment, marital adjustment, social adjustment, family adjustment, and personal adjustments. The semistructured interview starts in each major area with a non-directive lead and "priming" questions follow to insure coverage of each area. During the interview the subject is rated as to his adjustment on a series of continua corresponding to the priming questions. Qualitative data relevant to each of the ratings are also recorded. In addition, at the termination of the interview the interviewer records a score (from 0 to 9) which he feels would best reflect by rating the adjustment of the person interviewed. This is denoted as the *global estimate.*

During the interviews two kinds of ratings were recorded. One indicated how the interviewer felt the subject was perceiving himself (*e.g.*, with regard to social adjustment). In this case the interviewer estimated how the subject would rate himself had he the opportunity to do so. This is referred to as the *interviewer's estimate of the subject's self rating.* The sections containing these rating scales were compiled in a single form and administered to the subject after the interview proper. It provides a score—the *subject's self rating.* The object of this procedure is to assess understanding and empathy by comparing the *subjects' self rating* with the interviewers' estimate of that rating and thus provides a reliability check. Another section of ratings reflects the interviewer's *own* reaction to the subject. These are more in the form of summary judgments for each content area—and are denoted as the *interviewer's rating.*

The interview, then, contributes two criteria: the *interviewer's rating* and the *subject's self rating.* The *global estimate,* in substance, is similar to the *interviewer's rating* and serves the same function although it involves no continuous rating. The *interviewer's rating of the subject's self rating* is included as indicated as a reliability measure.

2. *The inventory criterion.* The adjustment inventory chosen as a criterion measure had to be capable of administration in a brief period *prior* to the *interview.* The morale scale of an instrument employed in studies by Rundquist and Sletto [5]

[1]The indices resulted from averaging standard scores on each of the instruments and restandardizing the averaged scores.

was revised for this purpose. The resulting 20 item instrument had not been revalidated prior to the present study, however.

3. *The community reputation criterion.* To obtain data concerning the subjects' social adjustment, some 50 adults in Nobles County communities who had familiarity with the subjects were interviewed. They ranged from gas station attendants to community leaders. Each judge was given a list of subjects' names and uniformly asked how he felt the subjects were "getting along". Responses were recorded by the interviewer and later content analyzed and scored. The scores were derived so that each judge was his own criterion for good and poor adjustment. This was accomplished by having the judge nominate the 3 best and 3 poorest among those he had previously described, and indicated the bases for the nominations. These provide the anchor points for scoring the individual reputations and serve in addition as community criteria of mental health.

Results

For the validation phase, four years subsequent to the original predictions, a sample of 76 males was interviewed and other data concerning them collected. The subjects are approximately equally distributed among the three (high, middle, low) *predictor* groups. Those interviewed were among those *not* in military service, moved from the county, or away at colleges too remote to be contacted (2).

In relating the predictors to the criteria, three kinds of analyses were undertaken: (1) calculation of the rank order correlation between predictors and criteria, (2) calculations of the percent hits between those *predicted* as high, middle, and low, and those judged as high, middle, and low according to the criteria, and (3) computation of the intercorrelations among the criteria. The results of the intercorrelations are presented in Table 1.

Table 1. Correlations (Rank Order) Among 1950 Predictors and 1954 Criteria of Adjustment

	2 Teacher Index	3 Comb. P & T Index	4 Interviewer's Rating	5 Global Estimate	6 Rundquist-Sletto Score	7 Reputation Score	8 Subject's Self Rating	9 Interviewer's Estimate of Self-Rating
Predictors								
1. Pupil Index	.47*	.81*	.35*	.31*	.15	.29**	.02	.29*
2. Teacher Index		.82*	.18	.14	.05	.41*	−.18	.13
3. †Combined Pupil Index and Teacher Index			.35*	.29**	.16	.32*	−.06	.27**
Criteria								
4. Interviewer's Rating				.91*	.38*	.31*	.49*	.70*
5. Global Estimate					.42*	.35*	.51*	.69*
6. Rundquist-Sletto Score						.26**	.39*	.37*
7. Reputation Score							.23	.37*
8. Subject's Self Rating								.61*

†Pupil Index and Teacher Index weighted equally
*Significant at .01 level
**Significant at .05 level

The relationship of the 1950 predictors—the Pupil Index (PI), the Teacher Index (TI), and *an arithmetic averaging of these* (PI + TI) to the criteria just outlined should provide some indication of the predictive value of these measures. The

results indicate the size of the correlations with the criteria range: for the Pupil *Index*, from .02 (*subject's self rating*) to .35 (*interviewer's rating*); the *Teacher Index* correlates from −.18 (*subject's self rating*) to .41 (community reputation score). The combined indices range from −.06 to .35. For *test* validation purposes these correlations though positive are relatively low. There is reason to believe, from other evidence, that having taken sample slices from our original population for the present study some correlations will be higher by about .20 for the sample than for the population. This fact, however, will have little effect upon the conclusions drawn because the size of the correlations obtained from the investigation are in the main already low.

From an *analysis of hits* in which the *percentage* of direct hits (*i.e., predicted* as high, middle, or low and judged 4 years later as high, middle, or low according to the particular criterion) was determined, one would conclude that the predictive value of the instruments using the present criteria is essentially as already indicated. With the *Pupil Index, Teacher Index* and *combination* of these as *predictors* and the *interview rating* as criterion there are from 41% to 47% hits. Against the *subject's self rating* the hits range from 29% to 33%; against the *Rundquist-Sletto revision* from 24% to 25%, and against the *community reputation scores* from 19% to 21%.

Discussion

We feel the explanation of these results should not be ascribed simply to the inadequacy of the instruments or the criteria, although this may be, at least in part, true. The possibility exists that the results are a consequence of changes in the attitudes and behaviors of the individuals about whom predictions were made. In the community reputation interviews, subjects have been described as very well adjusted as adults in contrast to what they had been in adolescence often enough to be skeptical of prediction measures which rest on a premise of consistency of behavior. Our observation is that a variety of behaviors characterize adolescents and children which for some will change, others not. To make more accurate predictions, then, from childhood or adolescence to adulthood will require more understanding of the developmental aspects of adjustment; in particular those characteristics which remain stable and those which change and under what circumstances.

A very different interpretation however is also possible, *i.e.*, that the "basic" personality structure, if known, would provide a stable and more adequate guide to prediction. The instruments in this study have not done as well as they might, the argument could go, because they have not adequately assessed this basic structure. This may be, but only from further empirical research will this be known. To date, studies in which predictions are based upon expected consistency of behavior from the time of prediction to the time of validation have been lacking in spectacular success.

An interesting relationship appears in the correlations of the predictive indices to the interview and community criteria. The *Pupil Index* (the composite of psychological instruments) predicts higher (.35) on the *interviewer's rating* (the psychologist's criterion) than on the community reputation criterion (.29). The *Teacher Index* (which includes nominations of best and least well adjusted, and personality ratings) conversely predicts higher on the *community reputation* criterion (.41) than on the psychologist's criterion (.18). In other words, for the men studied here, the psychologist's predictor predicts better against a psychologist's criterion and the non-psychologist's predictor predicts better on a non-psychologist's criterion. The combined indices predict equally on both criteria (.35, .32).

Of the present criteria, the subject's rating of his own adjustment (*the subject's self rating*) is predicted with the lowest correlations by both the Pupil and Teacher Indices (PI .02; TI −.18). The *community reputation* and *interviewer's ratings* are about equally well predicted, with the highest correlations obtained (PI *vs.* interviewer's rating: .35; TI *vs.* reputation score: .41). The Rundquist-Sletto instrument is somewhere between these (PI .15; TI .05).

It might appear from the correlation of .91 between the *interviewer's rating* and the *global rating* that being so highly correlated one could be discarded, probably the *interview rating* since it involves much more computation. There would be no hesitation in suggesting this, were it not that the *global rating* was made only after the interviewer had made the individual item ratings, though without knowledge of the total rating score.

The correlation of .61 between the *subject's self rating* and the *interviewer's estimate of the subject's self rating* can be regarded as reasonably high, considering the possible influences which might attenuate this relationship as, for example, the tendency for a subject to perceive the meaning of some items differently from the interviewer, and the occasional inconsistency between written and verbal production.

Summary and Conclusions

For the purpose of predicting adjustment of children and adolescents two series of instruments were developed utilizing self-ratings and teacher evaluations. Three predictive indices were computed from these instruments and related to criterion measures of adjustment (interview, inventory, self-ratings and community reputations) four years subsequent to the original testing. The data reported are for a selected sample of male adolescents who in the validation phase were young adults.

The correlations of predictive indices to validating criteria range from −.18 to .41, and in percentage "hits" from 19% to 47%. Considering the general lack of data concerning base rates in prediction of mental health (rather than mental disorder) it is difficult to adequately assess the success of the present predictions in light of the essentially "healthy" character of the present sample. The possible limitations of prediction studies based upon a premise of behavior consistency from the time of prediction to validation are discussed.

References

1. Beilin, H. The effect of social (occupational) role and age upon the criteria of mental health. *J. Social Psychol.* (In press).
2. Beilin, H. and Werner, E. The interviewing availability of a follow-up sample of rural youth. *Public Opin. Quart.* (In press).
3. Erikson, E. H. Growth and crises of the healthy personality. In *Symposium on the healthy personality.* New York: Josiah Macy Jr. Foundation, 1950.
4. Jahoda, M. Toward a social psychology of mental health. In *Symposium on the healthy personality.* New York: Josiah Macy Jr. Foundation, 1950.
5. Rundquist, E. A. and Sletto, R. F. *Personality in the depression.* Minneapolis: University of Minnesota, 1936.

FACTOR ANALYSIS OF RORSCHACH SCORES OF SCHIZOPHRENICS

BERNARD A. STOTSKY

Boston University

PROBLEM

Factor analysis of Rorschach scores has attracted a number of investigations concerned with analysis of conventional scoring categories [1, 4, 5, 6, 11, 12, 14, 15] though two have dealt with content [6, 13] and one with diagnostic signs [7]. Such studies have generally found a productivity factor. Wittenborn [14] using college undergraduates, discovered three additional factors: one heavily weighted on *W*, *K*, and *CF*; a second on *M*, *m*, *Fc*, *C'*, *FC*, *P*, and *O* on which *FC* was more highly related to *M* than to *CF* and *C*; and the third on *P*. His original findings were supported in a second study on a mixed group of psychiatric patients [15] and by Williams and Lawrence [11] on a different psychiatric population who also found a distinct intelligence and a shading factor. Combining the Rorschach and MMPI they again obtained a productivity factor and two factors, labelled "ego-strength" and "expressive-repressive", which showed overlap between the two tests [12]. Coan [4], analyzing twelve determinant scores from Wittenborn's original data, found seven factors none of which could be considered a productivity factor. He criticizes previous studies on two major grounds: (a) The productivity factor is an artifact resulting from the inclusion of total number of responses (*R*) as a scoring category. Since the correlations of *R* with any location or determinant scores are really part-whole correlations, *R* may be expected to correlate spuriously with many of the scores. Omission of *R* does not eliminate productivity entirely since other variables like *F* and *D* may be heavily weighted with this factor. It does, however, remove a source of spurious correlation.[1] (b) Derivation of location and determinant scores from the same set of responses is another source of spurious correlation and increases the experimental interdependence of certain categories.

A third criticism applicable to all but two studies [11, 12] is the questionable use of Pearson *r* in view of the short range of most Rorschach scores and the very high frequency of zero scores on most variables. Many variables have either curvilinear or point distributions. Tetrachoric *r* seems somewhat more appropriate for distributions which can be dichotomized, but assumes underlying continuity, normality, and linearity of regression. Categories with more than half of their subjects giving zero responses cannot meet these criteria. Most of the remaining categories for which less than half of the subjects give zero responses also show considerable skewness but come closer to meeting the assumptions for correlational techniques.

Consequently the following requirements were set up for the factor analysis of Rorschach data: (1) Elimination of *R* from the correlation matrix. (2) Treating location and determinant variables in separate matrices and omitting such content responses as animal, human, popular, and original responses. (3) Including only those variables for which more than half of the subjects gave one or more responses.

METHOD

Subjects. Although schizophrenics have been frequently studied by means of the Rorschach, none of the factor investigations has been on this diagnostic group. The author used Rorschach data from a previous study of 148 diagnosed schizophrenics in two VA neuropsychiatric hospitals. These schizophrenics were divided into two groups of 74 each. One group remitted and was later discharged. The second group was continuously hospitalized for two years after admission. Testing had been performed soon after admission.

Procedure. The protocols were scored according to Beck's earlier procedures [2]. Reliability of scoring had been established in the previous study [10]. Using the median as a cutting point, fourfold tables were set up for each pair of variables being

[1] Percentage scores are unusable because of their ipsative nature.

correlated: *W*, *D*, and *Dd* for location; *F*, *M*, *FSH*[2], *FC*, and *CF* for determinants. Tetrachoric *r*'s were then calculated from computing diagrams[3]. The resulting correlation matrices were factored by the Thurstone centroid method. Factoring was continued until a communality smaller than .25 was obtained for a factor. The centroid factors were then rotated orthogonally using the criteria for simple structure. These operations were performed first for the entire sample, then for both subsamples. A further analysis was made for a content, a determinant, and a location variable which the author felt shared high common variances, to serve as an illustration of the spurious correlation obtained when such variables obtained from a single set of responses are treated together in a correlation matrix.

Results

Location. Two factors were extracted. For all three groups Factor I′ is heavily loaded with *D* and *Dd*. Factor II′ tends to load on *W*, with zero loadings on *D* and *Dd* for the entire sample. Similar loadings are obtained for the subsamples, with the one for *W* reaching significance in the remitted group. Factorially the three groups are very similar.

Table 1. Intercorrelation of Rorschach Scores for Schizophrenics[1]

Scores	Total Sample						Remitted Schizophrenics						Hospitalized Schizophrenics					
	D	Dd	F	M	FC	FSH	D	Dd	F	M	FC	FSH	D	Dd	F	M	FC	FSH
W	−20	−08					−30	−34					−22	−16				
D		57	89					63	82					57	90			
M			42						30						42			
FC			29	43					41	31					14	20		
FSH			18	56	60				27	16	24				10	37	65	
CF			21	07	14	16			06	00	44	−24			13	16	02	12
P	62		53				69		62				46		51			

[1]Because of magnitude of S. E. of tetrachoric *r* and relatively small N, lower limit of .40 was set for significance of r's and factor loadings, as was done in a previous study[11].

Table 2. Rotated Factor Matrices for Rorschach Location Scores

Scores	Total Sample			Remitted Schizophrenics			Hospitalized Schizophrenics		
	I′	II′	h^2	I′	II′	h^2	I′	II′	h^2
W	-12	37	15	−39	41	32	−23	34	17
D	67	08	46	84	01	71	72	00	52
Dd	81	−01	66	77	−01	59	78	−04	61

Determinants. Two factors were extracted from the five determinants. For the total sample and for the hospitalized subsample one factor is most heavily loaded with *F* and *M* while the other is loaded for *FC* and *FSH*. For the remitted subsample one factor is loaded with *F*, *M*, *FC*, and *FSH*, the other with *FC* and *CF*.

Table 3. Rotated Factor Matrices for Rorschach Determinants

Scores	Total Sample			Remitted Schizophrenics			Hospitalized Schizophrenics		
	III′	IV′	h^2	III′	IV′	h^2	III′	IV′	h^2
F	62	11	40	03	64	41	−01	63	40
M	59	39	50	00	49	24	16	66	46
FC	39	62	54	62	57	71	76	16	60
FSH	23	79	68	−27	44	27	79	25	69
CF	40	−02	16	69	01	48	01	28	08

[2]Shading was treated as a single category in accordance with the procedure of the original study (10 p. 123 ff.)

TABLE 4. ROTATED FACTOR MATRICES FOR D, F, AND P

Scores	Total Sample			Remitted Schizophrenics			Hospitalized Schizophrenics			
	V′	VI′	h^2	V′	VI′	h^2	V′	VI′	VII′	h^2
D	18	93	90	16	90	84	11	30	92	95
F	00	96	92	01	91	83	−16	19	95	96
P	51	56	57	46	68	67	51	45	52	73

D, F, and P. Two factors were extracted for the entire sample and remitted subsample. One, accounting for three-fourths of the variance, loads heavily with *F* and *D* and significantly for *P*. The other has a significant loading only for *P*. For the hospitalized subsample there are three factors, one loading significantly with *F*, *D*, and *P*; the other two with *P*. *D* and *F* are so similar in their loadings and so high in their intercorrelations as to raise the question whether they are not almost identical variables.

DISCUSSION

With respect to location, Factor I′ is clearly a measure of the tendency to respond to the details of the blots. *D* and *Dd* are closely related, the latter probably representing an extension of the former for some subjects who in the course of giving a long protocol use rare details as well as common details. The tendency to use the whole blot in responding is quite distinct from an orientation to details, so that *W* has a negative loading on Factor I′ and a positive loading on II′ for which *D* and *Dd* show zero loadings. There is thus ample justification for treating whole responses separately from detail responses but not for treating common and rare detail responses separately on the basis of present findings.

For determinants the findings are not as clear. In contrast to previous findings *FC* is not as closely related to *M* factorially as it is to *FSH* in two groups and to *CF* in one. A review of factor loadings in previous studies shows that *FC* clusters with other variables besides *M*: in one[12] even with *CF*. Thus the relationship of *FC* with *M* is a function of a factor which is loaded with other variables besides these two. In two groups the relationship between *FC* and *FSH*, hitherto obscured by the division of shading into many subcategories, appears to represent a controlled reaction to the color aspects of the blots. One is to the brighter, the other to the darker hues. In the remitted group the clustering of *FC* and *CF* responses may be due to the presence of greater control in the *CF* responses of this subsample so that they more closely resemble *FC* responses. There may be qualitative differences between groups for *CF* responses. In this connection Phillips[8] found that schizophrenics who improve following shock treatment showed a higher color drive score than those who did not improve. His color drive score was derived from analysis of content without reference to determinants but a major portion of the color drive scores appeared to come from *CF* responses.

F and *M* cluster together consistently and may represent an orientation to the form aspects of the blots, in contrast to *FC*, *FSH*, and *CF* which involve the use of color to elaborate responses.

The results for *F*, *D*, and *P* point up the danger of treating separately components which arise from the same basic source. While *P* has independent loadings, *D* and *F* are very similar. This may be observed also in their high intercorrelations in other studies, never lower than .69. These two factors, and to a lesser extent *P*, seem to reflect a tendency to respond to the blots with simple form in a commonly observed detail. The *D* and *F* factors are present together in approximately forty to fifty percent of all Rorschach responses of all groups. The differences between groups and much of the variance attributable to distinctive traits will generally be found in the remaining responses. Treating *R*, *D*, and *F* as completely independent variables and then finding a high common variance may be a highly refined form of gilding the Rorschach lily.

On one point the author is in complete agreement with previous studies, *i.e.*, that Rorschach scoring categories need a thorough reworking. A scoring system should be developed to maximize the independence of the various categories and to produce scores which can be handled by conventional statistical procedures. If this means revision of both administration and scoring of the test, let us get to work on it.

Summary

Factor analyses were performed for Rorschach location and determinant scores on a sample of 148 schizophrenics and on subgroups within the sample of 74 remitted and 74 hospitalized schizophrenics. Three changes from the procedures of previous studies were made. Total number of responses (R) was omitted from the correlation matrices. Location and determinants were treated separately. Only variables for which more than half of the subjects gave one or more responses were included in this study.

For location two factors were found. Common and rare detail scores clustered together on one, while whole responses tended to load on the other factor. For determinants two factors were found. On one, pure form and human movement responses clustered together. On the other, shading and color responses showed high loadings though not as consistently as F and M. Separate factor analyses were performed for F, D, and P — determinant, location, and content scores which were believed to be highly interrelated. They were found to have a high common loading on one factor, especially F and D. This illustrates the problem of including such scores in a common matrix on the assumption that they are independent.

References

1. Adcock, C. J. A factorial approach to Rorschach interpretation. *J. gen. Psychol.*, 1951, 44, 261-272.
2. Beck, S. J. *Rorschach's test.* New York: Grune and Stratton, 1945.
3. Chesire, L., Saffir, M. and Thurstone, L. L. *Computing diagrams for the tetrachoric correlation coefficient.* Univ. of Chicago, 1951.
4. Coan, R. A factor analysis of Rorschach determinants. *J. proj. Tech.*, 1956, 20, 280-287.
5. Cox, S. M. A factorial study of the Rorschach responses of normal and maladjusted boys. *J. genet. Psychol.*, 1951, 79, 95-115.
6. Hsu, E. H. The Rorschach response and factor analysis. *J. gen. Psychol.*, 1947, 37, 129-138.
7. Hughes, R. M. A factor analysis of Rorschach diagnostic signs. *J. gen. Psychol.*, 1950, 43, 85-103.
8. Phillips, L. *Personality factors and prognosis in schizophrenia.* Unpublished doctor's dissertation, Univ. of Chicago, 1949.
9. Sen, A. A study of the Rorschach test. *Brit. J. Educ. Psychol., Statist. Sect.*, 1949, 19, 142-143.
10. Stotsky, B. A. *Factors in remission of schizophrenics.* Unpublished doctor's dissertation, Univ. of Michigan, 1951.
11. Williams, H. L., and Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, 17, 261-264.
12. Williams, H. L., and Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, 18, 193-197.
13. Wittenborn, J. R. A factor analysis of discrete responses to the Rorschach ink blots. *J. consult. Psychol.*, 1949, 13, 335-340.
14. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, 14, 261-267.
15. Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, 14, 469-472.

EFFECTS OF DIFFERENT METHODS OF STRUCTURING THE RORSCHACH INQUIRY UPON DETERMINANT SCORES[1]

STANLEY W. KLINGENSMITH

University of Pittsburgh[2]

PROBLEM

Most Rorschach workers indicate that it is essential for the examiner to avoid suggesting determinants to subjects during administration of the inquiry. Without experimental verification, it is apparently assumed that the greatest potential source of error lies in E's focusing S's attention in various ways to blot characteristics that he would not otherwise have considered as determinants of his responses, thus biasing his report during inquiry. The rationale may include Rapaport's[7] view that ". . . the issue is not what is *perceptible*, but what is *spontaneously perceived*—that is, what features of the perceptual mass initiate association processes spontaneously and influence their course." Siipola[8] points out that examiners employ widely different criteria in judging whether a given determinant should be scored. Levin[6] questioned whether the inquiry actually does reveal valid information regarding determinants involved in free association responses.

In addition to the more obvious effects of differences in inquiry methods, the importance of the problem is further emphasized by studies which show that a variety of subtle factors may affect the number of determinants elicited. For example, Gibby[3] and Gibby and Stotsky[4, 5] investigated the influence of type of inquiry, examiner differences, and "determinant shift". The purpose of the present study was to investigate the changes in determinant scores related to changes in level of test structure.

METHOD

Levels of Test Structure. Four levels of test structure were employed. Structuring refers to the instructions and aids used to guide subjects in reporting. Written instructions were provided. A description of the levels of test structure follows.[3]

1. *Free association.* In structuring the free association, general statements which indicated the nature of the task were provided. Instructions similar to those recommended by Beck[1] were used.

2. *Inquiry S1.* Instructions preceding this inquiry requested description and explanation of responses. After responses were freely elaborated, subjects were asked: "Is there anything else that gives that impression?" or its equivalent.

3. *Inquiry S2.* Instructions emphasized the need to explain in great detail *everything* involved in the perception of responses. Subjects were encouraged to report: "What *else* in addition to the shape, if anything, helps you to see what you do?"

4. *Inquiry S3.* In this inquiry brief definitions of determinants followed an extensive explanation of the nature of determinants and the need for detailed reporting. The printed definitions served as a check-list during administration of this inquiry. Subjects were asked to give the part of the definition which applied to each reported determinant, for example, which of the colors: red, green, black, gray, etc. was responsible for their reporting color as a determinant.

Determinants. Only the presence or absence of the following determinants was considered, independent of their combination with each other: chromatic color (C), achromatic color (C′) vista (V), human movement (M), animal movement (FM), inanimate movement (m), texture (T), and form (F). To simplify the task for subjects C and C′ were combined in a single definition of color, and M, FM and m were com-

---

[1]Adapted from a dissertation submitted in partial fulfillment of requirements for the Ph.D. degree, University of Pittsburgh, 1956. The author wishes to express his appreciation to Dr. Roy M. Hamlin for his guidance and counsel regarding the research on which this article is based. The assistance of Mr. Alvis M. Eskey in determining the reliability of scoring, and Mrs. Mary G. Zima in arranging for participation of the student nurses is also gratefully acknowledged.

[2]The author is now employed at the Mental Health Clinic, Leon County Health Department, Tallahassee, Florida.

[3]Instructions employed are presented verbatim in the original work, along with discussion of structuring used by experimenters who employed the group method of Rorschach administration.

bined in a single definition of movement. These were separated during scoring of the protocols, according to the card area employed, the content of the response, and— in inquiry S3— the part of the definition given by S.

Subjects. The experimental group comprised 40 student nurses, 20 in their second, and 20 in their third year of training (mean age 19.82). Base hospital affiliations throughout Greater Cleveland (Ohio) and several smaller communities indicate wide diversification of general background. A control group of 40 student nurses (mean age 20.05) was formed by matching for base hospital affiliation, year in training, and day of testing.

Test administration. All tests were individually administered by E. The procedure for the experimental group was as follows. The free association was administered to S in the usual manner, except that only two responses per card were permitted. Inasmuch as all subjects could not be expected to give two responses per card, fifteen of the 20 possible free association responses were selected by a predetermined method prior to administration of the first inquiry. Then inquiries S1, S2, and S3 were given in that order, one after the other. Efforts were made to copy responses verbatim.

The control group was given the free association and inquiry S3 only. The purpose of the control group was to determine the influence of practice effects and repeated exposure to the stimulus materials, *i.e.*, the cumulative effect of several inquiries upon the number of determinants reported by the experimental group in inquiry S3.

Treatment of the Data. Criteria for scoring were rigorously defined on the basis of other test records. In the free association and inquiries S1 and S2, determinants were scored on the basis of direct or indirect evidence in subjects' verbalizations that determinants were used. Basic content of certain responses was regarded as sufficient evidence. In inquiry S3 decision regarding use of determinants was left entirely to subjects rather than to E.

The free association and inquiries S1 and S2 of 20 protocols of the experimental group were scored independently by E and another Rorschach worker. Inter-scorer reliability is indicated by 94 per cent agreement for all determinants, and 89 percent agreement for determinants other than Form. E's scoring of the protocols was used in tabulating results.

Form was scored for most of the responses at each level of test structure. This determinant was therefore excluded from total score comparisons inasmuch as its inclusion would obscure differences among other determinants. Since distributions of scores for separate determinants were skewed, the nonparametric rank test and sign test [2] were applied to test the significance of differences. The "*t*" test was used in testing the significance of differences of combined determinant scores—total less Form.

Results and Interpretation

Table 1 shows a comparison of the experimental and control group scores in the free association and inquiry S3. Chance factors may easily account for the differ-

Table 1. Comparison of the Scores of the Experimental Group and the Control Group

Determinants	Free Association Group E	C	Diff.	*P*	Inquiry S3 Group E	C	Diff.	*P*
Chromatic color (C)	30	35	−5	.45	141	117	24	.10
Achromatic color (C')	9	3	6	.10	257	231	26	.43
Vista (V)	44	37	7	.86	196	191	5	.98
Human movement (M)	82	64	18	.25	105	82	23	.08
Animal movement (FM)	72	68	4	.62	220	214	6	.73
Inanimate movement (m)	26	15	11	.05	43	40	3	.83
Texture (T)	39	42	−3	.78	360	375	−15	.71
Total (Form excluded)	302	264	38	.23	1322	1250	72	.36

ences that occur. The data indicate that the increase in determinant scores in inquiry S3 for the experimental group is essentially independent of the effects of having taken inquiries S1 and S2.

Table 2 contains the scores of the experimental group for the free association and three levels of inquiry. There is an increase in number of determinants at each level of inquiry, except for m at inquiry S3. (This decrease in m is apparently an artifact resulting from its being frequently overlooked in inquiry S3 because the definition was combined with that of M and FM. The m scores are omitted in the following comparisons and discussion.) The high level of significance obtained for most of the comparisons indicates a sufficient degree of consistency in the data to conclude that the trends are real. Human movement is readily elicited in the free association, while other determinants are more readily elicited with varying degrees of increased structuring.

TABLE 2. DETERMINANT SCORES OF THE EXPERIMENTAL GROUP AT EACH LEVEL OF TEST STRUCTURE

Determinants	Free Association	Inquiry S1	Inquiry S2	Inquiry S3
Chromatic color (C)	30	87**	117**	114*
Achromatic color (C′)	9	68**	129**	257**
Vista (V)	44	70**	92**	196**
Human movement (M)	82	92	96	105
Animal movement (FM)	72	131**	163**	220**
Inanimate movement (m)	26	52*	62	43
Texture (T)	39	79**	132**	360**
Total	302	579**	791**	1322**

*Significant at the five per cent level.
**Significant at the one per cent level.

Note—Significance values are for the increase in scores over the scores at the previous levels.

Table 3 shows the pattern of scores more clearly. Using the number of determinants in inquiry S3 to represent 100 per cent, Table 3 contains the percentages of occurrence at each level of test structure. On the average, only 23 per cent of the determinants were elicited in the free association. This is understandable inasmuch as there was no request for elaboration of responses. That only an additional 21 per cent of the determinants were elicited in inquiry S1, when there was a request for elaboration, indicates that subjects of about average intelligence do not readily communicate information regarding characteristics of Rorschach cards that account for their responses. Introspection with the aid of definitions provided in inquiry S3 resulted in a marked increase in the number of determinants reported; 40 per cent of the total number were elicited in this inquiry.

TABLE 3. COMPARISON OF THE SCORES OF THE EXPERIMENTAL GROUP, EXPRESSED AS PERCENTAGES OF S3 SCORES "OCCURRING" AT EACH LEVEL FOR EACH DETERMINANT

Level of Structure	Determinants C	C′	V	M	FM	m*	T	Total
Free Association	21	03	22	78	33		11	23
Inquiry S1	41	23	14	10	27		11	21
Inquiry S2	21	24	11	04	15		15	16
Inquiry S3	17	50	53	08	25		63	40
Total percentages	100	100	100	100	100		100	100

*Inanimate movement (m) not included because percentage values are misleading.

It may be noted in Table 3 that the pattern of scores for C′, V, and T is different from that of other determinants. Only three per cent of the C′ determinants were given in the free association as compared with 78 per cent of the M determinants. Fifty per cent of the C′, V, and T determinants were not reported until inquiry S3.

Human movement is readily elicited, while FM and C fall between M and the shading responses— C′, V, and T— in this respect. Study of the protocols showed that the high proportion of M determinants elicited in the free association is not an artifact related to criteria employed in scoring.

Discussion

The present study indicates that the number and pattern of determinants elicited may vary greatly with different methods of inquiry. This finding has important implications regarding reliability, validity, and standardization of the Rorschach. For example, the fact that many determinants are not readily elicited permits the influence of extrinsic factors such as set of the subject and examiner differences, probably to a greater degree than is generally realized A need for reevaluation of the great concern with the danger of "suggesting" determinants is indicated. Evidence in this study that exposure of the experimental group to inquiry S1 and S2 did not affect the results in inquiry S3 suggests that we are dealing with a relatively stable phenomenon. It is quite possible that perceptions are sufficiently "set" at the moment of closure that direct questioning regarding determinants will not substantially alter the perceptions or bias the report of S.

Summary

The effect of different methods of Rorschach inquiry upon number of determinants scored was tested. The free association was considered as one level of structuring. The three inquiries varied in degree of structuring from one which provided little instruction or pressure for elaboration of responses to one which provided explanation and definition of determinants. The following results were obtained:

1. There is a corresponding increase in the number of determinants elicited as the directions to subjects for reporting determinants are more detailed and specific. Even after having been administered the free association and two separate inquiries, in the third highly structured inquiry there was a marked increase in the number of determinants elicited, excluding form—40 per cent of the total.

2. The pattern of determinant scores varies greatly according to the level of test structure. Human movement is readily elicited, even in the free association. The shading determinants are not readily elicited in the free association or in inquiries with little instruction or probing.

3. The increase in number of determinants elicited when definitions were provided was not a function of the experience of two previously-administered inquiries which did not contain definitions of determinants.

References

1. Beck, S. J. *Rorschach's Test, I. Basic Processes (2nd Ed.)* New York: Grune and Stratton, 1950.
2. Edwards, A. L. *Statistical Methods for the Behavioral Sciences.* New York: Rinehart, 1954.
3. Gibby, R. G. Examiner Influence on the Rorschach Inquiry. *J. consult. Psychol.*, 1952, 16, 449-455.
4. Gibby, R. G. and Stotsky, B. A. The Relation of Rorschach Free Association to Inquiry. *J. consult. Psychol.*, 1953, 17, 359-364.
5. Gibby, R. G. and Stotsky, B. A. Determinant Shift in Psychoneurotic and Psychotic Patients. *J. consult Psychol.*, 1954, 18, 267-270.
6. Levin, M. M. The Two Tests in the Rorschach. *J. proj. Tech.*, 1953, 17, 471-473.
7. Rapaport, D., Gill, M. and Schafer, R. *Diagnostic Psychological Testing (Vol. II).* Chicago: The Year Book Publishers, 1946.
8. Shipola, Elsa M. The Influence of Color on Reactions to Ink Blots. *J. Personal.*, 1950, 18, 358-382.

PSYCHOMETRIC SCALES FOR THE RORSCHACH POPULAR RESPONSE

DOROTHY PARK GRIFFIN

North Carolina State Board of Public Welfare
Division of Psychological Services

PROBLEM

The "popular" response, the most quantifiable of all Roraschach variables, has received little attention from experimenters. Hertz[5] gathered in 1938 a list of "populars" agreed upon by three or more scorers, among whom were Beck, Loosli-Usteri, Gardner, and Vernon. Not until 1950 when Beck's Spiegel study[2] appeared, were there any more than brief references to this variable. Then in 1951 Molish's article on the popular response in records of normals, neurotics, and schizophrenics[9] and in 1952 Friedman's study[4] were published.

The concept of popularity or communality[12, 13] has been interpreted as reflecting "ability to adjust to the practical requirements of life, to a sufficient participation in collective or common thinking" (Oberholzer, Piotrowski); "conformity in thinking processes"; "ability to adapt one's thinking to that of the common man, ability to experience communal thinking" (Molish); "F plus at its most concentrated", "cooperativeness, ability to participate in common thinking or wish, hence adaptability" (Beck). Recent studies of normals and schizophrenics in which P is shown to be a differentiating factor[9, 10] support somewhat these interpretations.

Antecedent, however, to any interpretation is the need for precise definition of the P factor which can be established objectively only on a statistical basis. A response thus may be defined as "popular" according to the frequency with which it occurs in a normal population. P responses are simply those most frequently given to a specific area of a blot. Rorschach suggested a criterion of "once in three records", and Oberholzer also used this criterion. Others have used a criterion of once in six, particularly for children's Rorschachs. Other factors, such as the size of the sample upon which frequencies are based, IQ, socio-economic level, type of culture, etc., of course, need to be controlled. If, as Beck says, the P factor provides "the most satisfying opportunity for quantitative method in Rorschach test procedure" [2], it should certainly be quantifiable by way of the rating scale technique, and this we have attempted to do.[1]

PROCEDURE

For the development of such a scale it is necessary only to assume that the P factor applies to the continuum which ranges from originality on the one hand to extreme communality, approaching universality, on the other. To provide an objective measure of this continuum, the author, in 1948, compiled a four-point psychometric scale from the lists of "popular" responses published by the three chief authorities in the field; Beck, Hertz, and Klopfer. Beck's later study of the Spiegel Mail Order House employees and Molish's comparison of their records with those of neurotics and schizophrenics provide additional data for a scale of normal populars. The scales presented in this article are a revision which incorporates these additional data.

In Beck's earlier list, as in those of Hertz and of Klopfer, the criterion was that of frequency, although the statistical basis was not elaborated upon. In fact, Hertz was the only one who presented data to substantiate the criterion of once in three records. The items in Klopfer's[7] list, admittedly "a selected sample", were quite vaguely defined for example as, "any creature with body in center and wings" for the W of Card I, and "any winged creature" for Card V.

In Beck's recent study, however, he uses three criteria as follows: 1. Operation Stimulus: the percentage of subjects selecting the area and giving this response must

---

[1]Developed under the inspiration of Joseph Zubin, in whose *Manual of Projective and Cognate Techniques*[15] the author's first version of the scale appeared.

be three times or more as great as for the next most frequent response; 2. Operation Group: fifteen per cent of the group must give the P response to this area; 3. Any area which is P must constitute a minimum of six per cent of the selections in its test figure.

With these more rigid (though still rather arbitrary) criteria the resulting list differs in several respects from the list published in Beck's 1944 text. There are five omissions, two of areas and three of responses to areas retained as popular, *i.e.*, D6 Card IV, "foot or shoe", W Card VII "clouds", "butterfly", "moth" for D3 Card III, and "octopus" for D1 Card X. There are four additional areas, besides two changes in the interpretation of areas, *i.e.*, D4 "human figure" and the interpretation of D3 in Card I as "a person's legs"; D4, "tree", "bush"; D1 interpreted now as "rodent, rat, mice, squirrel" as well as "bears", "beavers", "muskrats" in Card VIII; D2 "dog", D12 "sheep" in Card X. The total number of areas is actually increased from 20 to 23 and the number of concepts involved is increased to 82.[2]

Klopfer finds popular responses in ten areas on six cards, involving 25 different concepts. Hertz finds populars in 24 areas on nine cards, involving 64 concepts. Molish's results yield eleven areas and 18 concepts.

The normals in Molish's study were the carefully selected Spiegel group which represented the middle two-thirds of the general population in intelligence, age, and socio-economic level. Chi-square values were run to sift out the responses which occurred with significantly greater frequency among normals than among neurotics or schizophrenics. The areas and concepts selected by normals with significantly greater frequency than by neurotics or schizophrenics are considered essentially normal responses and are utilized in the present scale through Ratings Two to Five.

The following steps were selected as the basis of the scale. The intervals between steps are not assumed to be equal.

0. Original, not appearing in any published frequency table: Beck revised list (B), Hertz (H), Klopfer (K), Molish (M).
1. Tending towards originality; Occurring only once in 100 records. Listed by Hertz as original. (See samples)
2. Common but not popular. Listed in one only of the published frequency lists. Tendency toward popularity.
3. Quite popular. Listed in two only of the published frequency lists.
4. Very popular. Listed in three of the four published frequency lists.
5. Extremely popular, almost universal. Listed in the three well known published lists, also found by Molish to occur with significantly greater frequency in normals than with schizophrenics or neurotics.

FIG. 1. POPULARITY SCALES: GUIDE TO THE PSYCHOMETRIC SCORING OF THE POPULAR RESPONSE IN RORSCHACH RECORDS OF ADOLESCENTS AND ADULTS

Rating 0. *Original: not found in any frequency table.*

Rating 1. *Tending towards originality.*

CARD	LOCATION (Beck)	RESPONSE
I	W	human pyramid made up of acrobats
	W	North wind with his mouth open and blowing
	W	nightmare that you see in dreams
	W	bug in center, crawling into torn leaf
	W	mountain with reflection in water below
	D6	Russian Cossack with bushy beard
II	W	tooth because of its roots
	W	insect flattened out as if stepped on
	W	cross section of the brain
	D4	elves sitting back to back, cone shaped hats on
V	Dd22	man in the mountain

[2]Beck states that the number of P associations resulting from this study is 21. Nevertheless, a count reveals 23 different locations. He includes one which he admits does not satisfy his first criterion, but which is very close and was in the old list, i.e., "butterfly", "moth" D3 Card II.

III	V	W	animal, hands up as if trying to grab butterfly, tatooed on body
		W	cave with butterfly in middle
	V	D4	hat with feather or puff on side of it
IV		W	Indian buffalo god
		W	king surrounded by animals
		D4	person's arms, of a swimmer
V	V	W	peacock design
	V	W	hammock with two people lying in it
		W	a fur cape
		W	book ends
		D2	two persons
	V	D3	cobras about to strike
VI		W	trumpeteers with horns, back to back
		D7 & Dd26	snake's head and reflection
		D3	chimney with rays around it
VII	V	W	frog, bisected
	V	W	marionettes dancing with big heads on them
		D2	dog balancing something on his nose
		D5	forest with lighter landscape in back
VIII		W	decorated bowl
		W	wood carving with animals
		W	leaf with parts torn out
		D1	ornament for a tree
		D8	boat with sails
IX	V	W	woman gazing into mirror seeing reflection
	V	W	woman in costume of the 90's
		W	hat, jacket, dress on a rack
	V	Dd26	person's hand, thumb up, pointing
		D3	person's funny profile as in a cartoon
		D1	wind as seen in fairy books, flying
X		W	tapestry
		W	wall paper
		W	fancy monogram
	V	Dd24	mummy
		Dd33	child's head

Rating 2. *Common but not popular.*

Card	Location	Response
I	W	face (H)
		any creature with body in center and wings (not rated elsewhere) (K)
	D3	legs, lower part of body, person's legs and back, child's legs (B)
		human form (M)
	D4	human figure (B)
	Side fig.	person (H)
II	W	butterfly (H)
		dancers, clowns, jitterbugs (B)
	Side fig.	man, lady (H)
	D1	rabbit, bull, rhinoceros (K)
	D3	butterfly, moth (B)
III	W	dressed up animals, two human figures in a bending position (K), persons doing something (H)
	D1	characters, caricatures, dancers, jitterbugs, stage figures, marionettes, humans, people, waiters (B)
	D3	bow, (B) moth (H)
IV	W	bat, (H), ape, gorilla, bearskin rug, large furry animal (B).
	D6	foot, boot, shoe (H)
V	W	any winged creature (not rated elsewhere) (K)
	D1	leg (animal or human) (B)
VII	W	Persons (H)
	D2	Person (H)
	D4	butterfly, moth (H)
VIII	D1	any kind of four legged animal in motion (K), rodent, rat, squirrel, beaver, muskrat (B)
	D2	butterfly (H)
	D3	spine, fish bone (H), bone structure, vertebra skeleton (B)
	D4	bush (B)

IX	D1	person bending over doing something (H)
	D2	animals face or head (H)
	D3	persons, clowns, ghosts, gremlins, witches, mythological creatures (B)
	D9	tree (H)
X	D1	crustacean (B)
	D4	garden snake, tobacco worm, eel (K), serpents (H), rabbit's face (H), head of an animal with long ears or horns, as of a donkey or goat (K)

Rating 3. *Quite Popular.*

Card	Location	Response
I	W	bird (H, K)
II	W or side of fig.	humans, persons (B, H)
	D1	animal (H, K)
III	D3	butterfly, (H, K) bowtie, hair ribbon (B, K)
IV	W	animal skin, hide, fur, pelt (B, H)
V	W	bird (H, K)
VI	W or D1	fur, rug, pelt, (B, H)
VII	D1	face or head of person (B, H)
VIII	D1	bears, mice (B, M)
	D3	ribs, backbone (B, H)
	D4	tree (B, H)
IX	D3	human figure (B, M)
	D4	person's face (B, H)
X	D1	lobster (B, M), octopus (H, K)
	D2	dog (including special breeds) (B, M)
	D4	caterpillars, worms, seahorses (H, K)
	D5	rabbit (B, H)
	D12	sheep (B, M)

Rating 4. *Very popular.*

Card	Location	Response
V	W	bat, butterfly, moth (B, H, K)
IX	D4	human head (B, H, M)
X	D5	rabbit's head (B, H, K)

Rating 5. *Extremely popular, almost universal.*

Card	Location	Response
I	W	bat, butterfly, moth
II	D1	dog, bear
III	D1	persons
VI	W or D1	animal skin
VIII	D1	four-legged animals
X	D1	crab, spider

Discussion

It should be noted that "response", as referred to in this scale of the popularity continuum, refers not merely to location. Neither is it solely a matter of the words used. Since different meanings are denoted by different words, all alternative verbalizations listed by the different authors are counted as different responses unless the words are essentially identical in meaning, *e.g.*, "persons" or "humans". This would seem to be the only accurate method of compilation, since some authors, (Beck, for instance) list many similar concepts as separate populars and yet expressly mention other similar ones which are not to be considered popular.

It is interesting to note that the most universal "popular" is still "bat", Card 1 W. In Hertz's 1938 article, she reported three areas and eight concepts upon which all scorers agreed. Besides the "bat" of Card I, the W of Card III, interpreted as, "persons, men, or clowns" and the W of Card VI, interpreted as "animal, skin, hide, fur, pelt" were reported by all investigators. These responses continue to rate high on the scale, as is seen in Rating 5. The W of Card V, "bat", is also highly popular. although not appearing in the Molish list. However, it is surpassed in frequency by the "bat" of Card I.

According to Ranzoni[11], Cards I, V, VI, VIII, and X commonly yield animal responses and Cards III and VII, human responses. The four lists used as a basis for our scale give animals for Card II also, while Hertz lists an animal or insect response for every card. Beck lists no animal response for Cards III, VII, or IX; Klopfer, none for Cards IV, VII, or IX; Molish, none for Cards III, IV, V, VII, or IX. Hertz and Beck list human responses for Cards III and VII, as in the Ranzoni study, and also for Cards I and II. Klopfer finds a human response popular only in Card III. Molish lists human populars in Cards I, II, and IX.

Bochner and Halpern quote a "list of the usually accepted popular answers" [3, p. 15] which appears to be based upon Klopfer's list with a few Beck populars added. It covers 11 areas and gives 16 different concepts. In another section of the book the authors refer to 18 popular responses in all [p. 68], but this writer counts only 16. No P are listed for Cards VII and IX. No statistical basis is given and no explanation for the inclusion or exclusion of any item or area.

McCall and Doleys[8] listed no specific popular responses, but on the basis of further analysis of Beck's Spiegel sample, constructed tables of rough norms for popularity saturation at various response levels. It is likely that psychometric ratings of P will also ultimately have to be corrected for varying levels of R.

Summary

A five step psychometric scale, based upon the lists of populars published by Beck, Hertz, Klopfer, and Molish, is presented as a tool for more refined, objective scoring of the P factor in Rorschach records of adolescents and adults. The present scale attempts to achieve greater objectivity and refinement which can be further tested both clinically and experimentally.

References

1. Beck, Samuel J. *Rorschach's test. Vol. 3. Advances in interpretation.* New York: Grune & Stratton, 1952.
2. Beck, Samuel J., Rabin, Albert, I., Thiesen, Warren G., Molish, Herman and Thetford, William N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, 30, 241-298.
3. Bochner, Ruth and Halpern, Florence. *The clinical application of the Rorschach test.* New York: Grune & Stratton, 1945.
4. Friedman, Howard. A comparison of a group of hebephrenic and catatonic schizophrenics with two groups of normal adults by means of certain variables of the Rorschach test. *J. proj. Tech.*, 1952, 16, 352-360.
5. Hertz, Marguerite. The popular response factor in the Rorschach scoring. *J.Psychol.*,1938, 6, 3-31.
6. Hertz, Marguerite. *Frequency tables to be used in scoring responses to the Rorschach ink-blot test.* Cleveland, Ohio: Western Reserve University, (Rev. Ed.) 1946.
7. Klopfer, Bruno and Kelly, Douglas M. *The Rorschach technique.* New York: World Book Co., 1942.
8. McCall, Raymond J. and Doleys, Ernest J. Jr. Popular responses on the Rorschach test in relation to the number of responses. *J. clin. Psych.*, 1955, 11, 300-302.
9. Moli h, H. B. The popular response in Rorschach records of normals, neurotics, and schizophrenics. *Amer. J. Orthopsychiat.*, 1951, 21, 523-531.
10. Molish, Herman B., Molish, Ellen E. and Thomas, Caroline B. A Rorschach study of a group of medical students. *Psychiat. quart.*, 1950, 24, 744-774.
11. Ranzoni, J. H., Grant, M. S. and Ives, V. Rorschach 'card-pull' in a normal adolescent population. *J. proj. Tech.*, 1950, 14, 107-133.
12. Zubin, J. and Young, K. M. Manual of projective and cognate techniques. Madison: College Typing Co., 1948.
13. Zubin, J. *Experimental Abnormal Psychology.* (mimeographed) New York: Columbia University Bookstore, 1953.

SELF CONGRUENCE, BY SEX AND BETWEEN THE SEXES, IN A "NORMAL" POPULATION[1,2]

JOHN G. MARTIRE AND ROBERT H. HORNBERGER

Southern Illinois University *State University of Iowa*

PROBLEM

The purpose of this study was to investigate for two unselected samples of presumably "normal" men and women, the hypothesis that positive, significant intrasex and intersex relationships would be found among three specific measures of the self. These measures were in the form of self ratings and were designed to assess a self-Ideal (Scale sI), an Actual self (Scale As) and the Social self (Scale Ss). The expectations of overall and specific positive relationships were made on the basis of theory, empirical data [1, 2, 3, 4, 5], and on parsimony. Specific separate analyses were computed for each sex sample and between the sexes on each self-rating scale. The obtained data may be seen in the context of self congruence theory as it applies to men, to women and to the *relationships between the sexes.*

PROCEDURE

Subjects. *Ss* were members of a pool of 76 male and 66 female, presumably well-adjusted, undergraduate volunteers who were obtained individually and in groups from university dormitories and from fraternities. From this larger pool, 40 men and 40 women were selected so that *Ss* in each sex sample (Table 1) would be similar in age, class distribution and grade point, to the opposite sex. Although comparable, the samples were not representative of the university population. Data for this selection procedure were obtained from a personal data blank completed by each *S*.

TABLE 1. CHARACTERISTICS OF THE SAMPLE

	Men		Women	
Factors	M	σ	M	σ
Age	19.3	2.05	19.4	1.41
Class Distribution	Sex samples nearly identical.			
Grade Point	2.63	.53	2.78	.46

Method. *Ss* met in small groups and were told that *E* was investigating and trying to understand the general attitudes of students in order to be of help to them. Their cooperation and attention to instructions were encouraged. Each *S* was given a folder which contained three forms of the self-rating scale (see below), a personal data blank and definitions of the twenty-six traits[3] to be rank-ordered. Two anchoring definitions were provided, indicating the extremes of each trait continuum which was to be utilized in making the self-ratings. For example, the anchoring definitions of the trait of sociability were: "A high degree of sociability means that the individual is friendly and interested in others and seeks contact with them; a low degree of sociability means that the individual tends to withdraw from others and keep to himself." *Ss* were instructed to read through the trait definitions and to then make

[1]The authors wish to express their gratitude to Dr. Harold P. Bechtoldt for his invaluable assistance in the statistical and editorial preparation of this manuscript. Dr. Leonard D. Goodstein made several helpful suggestions.

[2]The authors are indebted to Drs. Charles D. Spielberger and Leopold O. Walder and Miss Jacolyn Young who collected the data for this study.

[3]Autonomy, Creativeness, Dependability, Dominance, Emotional Maturity, Flexibility, General Success, Individuality, Initiative, Intelligence, Interest in Opposite Sex, Motivation, Overall Adjustment, Physical Attractiveness, Popularity, Prestige, Security Minded, Self Acceptance, Sincerity, Sociability, Self Confidence, Social Poise, Sportsmanship, Self Understanding, Trustfulness, Understanding of Others.

the self-ratings on the three forms, following closely the specific directions given on each form. They were told to make the self-ratings in a standard sequence and to keep the sets of ratings independent of each other.

For the self-Ideal (Scale sI), *S*s were instructed to rank the twenty-six items in order of *importance* to them. Rank 1 was to be given to that trait about which they were most concerned and in which it was most important to stand personally high on. A rank of 2 was to be given to the next most important trait and so on through number twenty-six. Number twenty-six, it was then pointed out, would be the trait of *least* importance to *S*s to stand high on.

The Actual self (Scale As) instructions were similar to those of Scale sI except that in this case, the rank of 1 was to be given to that trait which was most *characteristic* of the individual, *i.e.*, the trait which *S* felt was most typical of himself and most necessary for an accurate description and full picture of himself. It was explained that a large amount of a particular trait was not necessary in order for it to be characteristic of a person.

For the Social self (Ss), *S*s were instructed to assign the first rank to that trait which they considered most *acceptable* for a person in our society to have, *i.e.*, the trait which most people would consider a good one to have. The additional instructions were similar to Scales sI and As.

The personal data blank was administered last and was used to obtain, among other things, *S*s' ages, academic classifications and grade point averages, which were utilized in the subject selection process mentioned above. Within group comparisons were made for each sex sample by computing all possible intercorrelations[4] among the three scales. The item values used in these computations were mean ratings (N = 40) on each trait.

Results and Discussion

It can be seen in Table 2 that high, positive and significant intercorrelations exist among the three scales for men, indicating that these *S*s have similar conceptions of self as defined. Such an interpretation must be based on the assumption

Table 2. Within Group Correlations Between Each Pair of Scales for the Two Sexes and the Differences Between Correlations (N = 26 for each Cell)

Scales	Men		Women	
	r	Diff.	*r*	Diff.
sI-As	.84**		.87**	
		.11		.35**
sI-Ss	.73**		.52*	
		.04		.08
As-Ss	.69**		.44*	
sI As-As Ss		.15		.43**

** $p < .01$ * $p < .05$

that *S*s did in reality react differentially to the instructions. The fact that the differences between these correlations are not significant supports the aforementioned finding that there exists a high level of agreement among the three measures of the self, and supports the theoretical argument that "normal" adjustment and congruent self conceptions are positively related. It confirms, with a group of older *S*s, the positive relationship between the actual self concept and the ideal self concept reported by Hanlon *et. al.* (2) for a group of high school students.

The intercorrelations among the three scales for the women are all positive and significant. Reference to Table 2 shows that the sI-As correlation is significantly

[4]In all cases, the correlations were Pearson *r*'s.

larger than the correlations for sI-Ss and As-Ss. With these female subjects, the introduction of the Social self scale significantly lowers the levels of agreement or congruence among these measures of the self[5]. This obtained lack of congruence for the female *S*s implies that these women see themselves as being and behaving in a manner different from what they believe society expects of them. Both their hopes about themselves and their actual characteristics seem to be at variance with their understanding of social expectations. In dynamic terms, one may infer conflict between personal strivings and either present social standards or introjected values.

Table 3 shows the intersex correlations for each of the three scales. All of the coefficients are high, positive and significant, suggesting a high level of agreement between the men and the women under this set of instructions. The inference is made that the sexes have highly similar concepts of self, ideally, realistically, and socially. Again, reference to Table 3 shows that the Social self (Ss) correlation of .85

TABLE 3. CORRELATIONS BETWEEN MEN AND WOMEN FOR EACH SCALE AND THE DIFFERENCES BETWEEN THESE CORRELATIONS (N = 26 FOR EACH CELL)

Scales	*r*	Diff.
sI-sI	.93**	
		.06
As-As	.87**	
		.02
Ss-Ss	.85**	
sI sI - Ss Ss		.08*

** $p < .01$ * $p < .05$

is significantly lower than the self-Ideal (sI) correlation of .93, meaning that these men and women are more alike in terms of their self-Ideal than in terms of their respective social expectations. Such a finding may be attributed to different standards set for each sex.

In this context, although intersex comparisons between the correlations on each pair of scales (Table 2) show substantial absolute differences in two out of three cases (Diff. sI Ss_{M-W} = .21; Diff. As Ss_{M-W} = .25), none of these differences is significant. This certainly suggests, however, that women show lower self congruence than do men perhaps because these college women have greater social pressure to conform and are consequently less independent.

In order to ascertain the degree of subject agreement in the trait rankings, coefficients of concordance (W) were computed. The results of this comparison for the men, shown in Table 4, indicate that male *S*s ranked the traits similarly on the three scales. However, this level of agreement (.82) is taken to mean that while there was some consistency in the rankings, the coefficients were not so high as to preclude differential effects of the instructions which defined each scale. For the

TABLE 4. W VALUES FOR SCALES sI, As, and Ss, AS A GROUP AND INDIVIDUALLY FOR THE TWO SEXES

Scales	W Coefficients** Men	Women
sI, As, Ss	.82	.73
sI	.21	.37
As	.13	.19
Ss	.22	.25

**$p > .01$ in each case

[5]An additional analysis was made to ascertain which traits seemed to account for the significantly lower correlations (sI-Ss; As-Ss) for the women. When these traits were identified, their meaning seemed too confused and ambiguous to report here.

women, too, the order of the W coefficient (.73) allows for rejection of the hypothesis of parallel forms.

The coefficients obtained for each scale, for each sex independently, are low, although significantly different from zero, indicating considerable variability in the values assigned to any one trait. The women were consistently less variable than the men in their ranking of the traits on all three scales individually.

Summary

On the basis of theory and empirical studies, it was expected that positive intra-sex and intersex relationships would be found among three specific measures of the self in a presumably "normal" population of men and women.

Three scales, defined by their instructions as a self-Ideal, an Actual self and a Social self, were administered to unselected male and female samples of a university population. In general, the results showed high and significant correlations among the three scales for the men, meaning that these *S*s, assumed to be normal, do have, as expected, similar or congruent conceptions of self. For the women, the correlations are somewhat lower although still significant. It was found that the Social self scale introduced a factor which significantly lowered the levels of agreement between it and the other two scales for the women. This finding was interpreted to mean that these female *S*s see themselves as being and behaving in a manner different from what they believe society expects of them. This discrepancy was characterized as a conflict based on either "reality" or "distortion".

On the basis of the obtained high, positive and significant intersex correlations, it was inferred that the sexes have highly similar concepts of self, ideally, realistically, and socially. Self concept theory may only be applicable to within group congruence or discrepancy. These findings may be seen as some support for the self theory where it concerns men. There is doubt cast on either the specific adjustment level of these women or on the accuracy of the theory.

References

1. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, *47*, 597-606.
2. Hanlon, T. E., Hofstaetter, P. R. and O'Connor, J. P. Congruence of self and ideal self in relation to personality adjustment. *J. consult. Psychol.*, 1954, *18*, 215-218.
3. Martire, J. G. Relationships between the self concept and differences in the strength and generality of achievement motivation. *J. Personal.*, 1956, *24*, 364-375.
4. McKenna, Sr. H. V., Hofstaetter, J. C. and O'Connor, J. P. The concepts of the ideal self and of the friend. *J. Personal.*, 1956, *24*, 262-271.
5. Rogers, C. R. and Dymond, Rosalind F., Eds. *Psychotherapy and personality change.* Chicago: Univ. of Chicago Press, 1954.

SIMILARITY OF PERSONALITY IN THE SOCIOMETRIC PREFERENCES OF MENTAL PATIENTS[1]

WILLIAM G. SHIPMAN

University of Pittsburgh Medical School

PROBLEM

The concept of measuring a person's strong, inner drives by the study of his likes and dislikes for others known to have these same trends was used as the basis of the Szondi Test [3], but attempts to validate this principle using the test's unclear pictures of mental patients in outmoded dress have been generally unsuccessful [1]. This study attempted to bypass the problem of pictorial representation by using actual people with mental disorders—patients on a ward of a mental hospital. Not only were they the people being chosen but they were the choosers. Despite early notions to the contrary, psychotic patients have recently been shown to have measurable feelings for each other [5, 8] and there is a current growing awareness of and interest in the social structure among the mental patients on a mental hospital ward [10, 11, 12].

As in many of these previous ward studies the sociometric technique developed by Moreno [7] was used in the present study to discover whether the choices of these people with extremes of personality organization were significantly related to their own personality type. The hypothesis to be tested is that people of a given personality structure will tend to prefer as friends people of a similar personality structure.

METHOD

Subjects. The subjects were 59 mental patients. All of them were on one mental hospital ward of 115 male patients of fairly good adjustment. The remainder of the ward population was not tested because the patients were either too sick to cooperate or unreachable due to a tight schedule of work and recreational activities. Of the 59 cooperating subjects, 22 were considered mainly paranoid, 18 were chiefly schizoid, 7 were sociopathic, 4 were emotionally unstable, and 8 were unclassifiable. Twenty were rated as being in good psychological adjustment, 27 as poor, and 12 as very poor. The mean age was 43 and the mean education was tenth grade. There were no statistically significant differences between the paranoid and schizoid groups for age, education or rated psychological adjustment.

Procedure. Each patient was seen by E in a private room, reassured about the confidentiality of the material, then asked: "Suppose you have been given a free trip around the world, and you can take one of the men from the ward with you. Which one would you choose?" The patient's answer and explanation for it were followed by a similar inquiry about his second and third choices. The case folders were then consulted for the patient's age and education. The ward psychiatrist, who knew all the men well, estimated their level of psychological adjustment (good-poor-very poor) and classified their personality structure according to the following scheme which is drawn from the *A. P. A. Diagnostic and Statistical Manual* [13, pp. 34-38]:

1. Paranoid
2. Schizoid
3. Compulsive
4. Inadequate
5. Sociopathic
6. Emotionally unstable
7. Cyclothymic
8. Unclassifiable

The paranoid and schizoid categories were used to help differentiate among the paranoid schizophrenics who were either predominately schizoid or predominately paranoid.

RESULTS AND DISCUSSION

The hypothesis that people prefer others of a similar character structure was tested in two comparisons because the ward population was comprised predominately

[1]The author wishes to thank Doctors Schilling, Greizman, and Lane as well as the patients at Torrance (Penna.) State Hospital, for their cooperation in this study. The writer was on the hospital staff when this experiment was done.

of two groups, 45 paranoids and 37 schizoids. Comparisons of the degree to which these two groups tended to prefer others of a similar personality structure are presented in Tables 1 and 2. Table 1 reveals a much stronger than chance tendency for

TABLE 1. EXPECTED AND OBTAINED PREFERENCES OF PARANOIDS

Item	Paranoid	Schizoid	Others	Total
Frequency of patients in ward	45.0	37.0	33.0	115.0
Expected chance ratio of patients' *total* choices	19.7**	16.5	14.8	51.0
Obtained *total* choices by patients*	36.0	5.0	10.0	51.0
Expected chance ratio of patients' *first* choices	7.7	6.5	5.8	20.0
Obtained *first* choices by patients†	13.0	1.0	6.0	20.0

*X^2 is 29.42, $p < .01$.
†Cell frequencies are too small to calculate X^2.
**Since a patient could not choose himself the expected frequencies for the paranoids are $N - 1$.

the paranoids to prefer other paranoids while Table 2 shows the preferences of the schizoids to be distributed in an apparently chance manner. The rationale of the Szondi Test states that the category chosen indicates the extent of the particular need tension. Since there is a markedly similar need tension system in all patients of the same personality structure and a generally similar degree of acceptance of it there should be strong mutual preferences. The behavior of the paranoids seems to fit this theory but the behavior of the schizoids seems a serious challenge to it.

TABLE 2. EXPECTED AND OBTAINED PREFERENCES OF SCHIZOIDS

Item	Schizoid	Paranoid	Others	Total
Frequency of patients in wards	37.0	45.0	33.0	115.0
Expected chance ratio of patients' *total* choices	11.7**	14.6	10.7	37.0
Obtained *total* choices by patients*	12.0	15.0	10.0	37.0
Expected chance ratio of patients' *first* choices	5.7	7.1	5.2	18.0
Obtained *first* choices by patients†	6.0	9.0	3.0	18.0

*X^2 is .0078, $p > .99$.
†Cell Frequencies are too small to calculate X^2.
**Since a patient could not choose himself the expected frequencies for the schizoids are $N - 1$.

Apparently the paranoid patients responded positively to themselves in others while the schizoid patients did not even react negatively to it. This suggests that the paranoids were more personally involved in their social milieu than were the schizoids. This is consistent with general concepts about the two conditions [9, 13]. A schizoid commonly avoids close relations with others and tends toward autistic thinking while paranoids are characterized by marked sensitivity in interpersonal relations and extensive use of the projection mechanism. The restitutional nature of paranoid symptoms is considered to keep the patient functioning well while his thinking is psychotic about certain specific matters [4].

Most experts consulted about this project expected the paranoids to dislike each other because they would sense and avoid each other's suspiciousness and unconscious homosexual impulses[4, p. 427]. Psychological research also suggests this in that Cohen[2] found people who project to be the most mutually antagonistic of all the groups he studied. Perhaps this strong mutual preference of the paranoids was facilitated by these unconscious factors as well as by a possible readiness to believe and support each other's delusional character defenses.

The apparently random preferences of the schizoids raises the question of whether they did, indeed, choose on a random basis or whether it was done in some personally meaningful way that was independent of perception of self in others. In order to answer this question the preferences of the two patient groups were studied to see whether the factors of similarity of age and ward location were operative in a differential manner. It was thought that they would operate more strongly on superficial choices than on choices determined by more personal factors. The obtained Pearson correlations are all small and statistically insignificant except for one for the paranoids ($r = .29$) which indicates a mild tendency for them to choose as friends people living near their quarters on the hospital ward.

The reasons given by the patients for their preferences were searched for clues as to the basis of the schizoid's selections. Aside from the greater number of reasons and the greater specificity in the paranoid's answers there were no differences between the groups. These further checks do not seem to help answer the question of what determinants led to the choices of the schizoids. All that we can conclude is that the choices were made in an apparently random manner despite the fact that the schizoids cooperated with the study and gave meaningful but undifferentiated reasons for their choices.

Summary and Conclusions

An attempt was made to test the tenet of Szondi Test theory that claims mutual preferences will appear amongst people of similar personality structure. This was done by studying the sociometric preferences of 38 mental patients, 20 of whom were considered to have a paranoid personality structure and 18 who were considered to have a schizoid personality structure. The paranoids had a statistically very significant preference for each other while the schizoids chose each other only to a degree expected by chance. The behavior of the paranoids was interpreted as lending support to the hypothesis while the behavior of the schizoids was felt to be a serious challenge. It could not be determined whether the unpredicted behavior of the schizoids was due to complete emotional withdrawal or to relating to others in a manner independent of one's perception of self in others.

References

1. Borstelmann, L. J. and Klopfer, W. G. The Szondi Test: A review and critical evaluation. *Psychol. Bull.*, 1953, *50*, 112-132.
2. Cohen, A. R. Experimental effects of ego-defense preference on interpersonal relations. *J. abnorm. soc. Psychol.*, 1956, *52*, 19-27.
3. Deri, Susan. *Introduction to the Szondi Test, theory and practice.* New York: Grune and Stratton, 1949.
4. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
5. Kegeles, S. Hyde, R. W. and Greenblatt, M. Sociometric network on an acute psychiatric ward. *Group Psychother.*, 1952, *5*, 91-110.
6. McMillan, J. J. and Silverburg, J. Sociometric choice patterns in hospital ward groups with varying degrees of interpersonal disturbances. *J. abnorm. soc. Psychol.* 1955, *50*, 168-172.
7. Moreno, J. L. *Who shall survive?* Washington: Nervous and Mental Disease Pub. Co., 1934.
8. Pope, B. Sociometric structure and group psychotherapy in a mental hospital for criminally insane. *Group Psychother.*, 1953, *5*, 183.
9. Reichmann, F. F. *Principles of intensive psychotherapy.* Chicago: University of Chicago Press, 1950.
10. Rowland, H. Friendship patterns in a state mental hospital. *Psychiatry*, 1939, *2*, 363-373.
11. Schwartz, M. S., and Stanton, A. H. A social psychological study of incontinence. *Psychiatry*, 1950, *13*, 399-416.
12. Schwartz, M. S. and Will, G. T. Low morale and mutual withdrawal on a mental hospital ward. *Psychiatry*, 1953, *16*, 337-353.
13. *A. P. A. Diagnostic and Statistical Manual.* Washington: Amer. Psychiat. Assoc., 1952.

SELF AND PEER PERSONALITY RATINGS OF PSYCHOTHERAPISTS

ELI A. RUBINSTEIN AND MAURICE LORR*

Veterans Administration and Catholic University of America

Problem

Studies devoted to objective measurement of the personality characteristics of persons in various professions dealing with mental illness have been relatively few [4]. And within this area no work seems to have been done in providing any comparative personality descriptions of psychiatrists, clinical psychologists, and psychiatric social workers to discover any distinctive group characteristics. In view of the present acceptance of the "team" approach to psychiatric treatment, especially in outpatient clinics, it would also be useful to know if any significant group differences in personality exist among these three disciplines.

As part of a larger research on evaluation of change during psychotherapy we attempted to obtain some information on the personality characteristics of therapists and how those characteristics may be related to patient changes in treatment. The various clinics collaborating in the major study were invited to send in personality ratings of clinic therapists, including psychiatrists, psychologists, and social workers who were all seeing patients for intensive treatment. This original purpose of this phase of the research could not be carried out because not enough of these ratings were on those particular therapists who were working with the patients in our research study. There are nevertheless, some interesting findings that can be reported on some of the characteristics of therapists as revealed in these data.

Procedure

The Multidimensional Scale for Rating Psychiatric Patients, previously described [3], was completed by 19 psychiatrists, 27 psychologists, and 21 social workers. These were all full-time VA personnel in various outpatient clinics.[1] All functioned as therapists in outpatient treatment at the time of this study. These 67 persons rated themselves on the rating scale. At the same time, two peers of the same or other disciplines who were well acquainted with the subjects also independently rated them on the MSRPP. In some cases a rater described more than one colleague. In other instances, the person rated also acted as a rater. The total number of individuals involved was 86. All ratings were kept anonymous and were coded to allow for matching of self and peer ratings and to designate the discipline of the ratee and the rater.

The analysis of the ratings was in terms of factors previously isolated [3]. Six of these factors are essentially descriptive of dynamic personality characteristics while four are mainly in terms of symptoms. The analysis of the therapists' ratings was restricted to these six dynamic factors. A factor score consisted of the mean rating on the group of items which best defined a factor. In all, 25 items were included in the rating.

Results and Discussion

Table 1 shows the mean factor scores by discipline and by factor for all subjects and for both self and peer ratings. The possible range of scores is from 1 to 6, with 3.5 a mid-point score. In no case is there a significant difference, in terms of the usual "t" test, between mean ratings for the three disciplines on any single factor.

---

*From the Veterans Benefits Office, Washington, D. C. The authors gratefully acknowledge the assistance of Elizabeth Turk with the statistical computations.

[1]We are indebted to the VA Mental Hygiene Clinics and the respective clinic staffs in the following cities for cooperating in the collection of this data: Baltimore, Brooklyn, Chicago, Denver, Newark, Pittsburgh, St. Louis, St. Paul, San Francisco, and Washington, D. C.

TABLE 1. MEAN FACTOR SCORES OF THERAPISTS

Ratings	Factors A	B	C	D	E	F
Self						
Psychologists	3.48	3.04	2.54	3.40	4.36	4.33
Psychiatrists	3.55	2.86	2.46	3.52	4.35	4.21
Soc. Worker	3.49	3.08	2.63	3.31	4.41	4.31
Peer (1)						
Psychologist	3.64	3.12	2.56	3.24	4.18	4.45
Psychiatrist	3.42	3.39	2.92	3.48	3.93	4.13
Soc. Worker	3.66	3.46	2.84	3.40	4.07	4.44
Peer (2)						
Psychologist	3.52	3.21	2.64	3.29	4.24	4.45
Psychiatrist	3.43	3.37	2.89	3.59	4.02	4.38
Soc. Worker	3.54	3.19	2.71	3.51	4.11	4.25

Nor is there in any case a significant difference between self and peer rating for a particular discipline on a particular factor. Briefly, the six factors may be described as follows:

Factor A involves emotional responsiveness. At the high end, in terms of our scoring, is the emotionally restrained person who rarely expresses his feelings overtly. At the other extreme is the uninhibited individual who easily expresses his feelings.

Factor B measures suspicious rebelliousness and a lack of concern for others versus trusting acceptance of authority plus much concern for the welfare of others. A high score indicates the former end of the continuum.

Factor C defines the degree of manifest tension and irritability. The higher the score, the greater the tension.

Factor D is concerned with the sense of personal adequacy. At the high score end the individual has a weak belief in himself, gives way to others and usually feels inadequate. At the other extreme there is a very strong feeling of personal adequacy.

Factor E is descriptive of motivation for long term goals. A high score is indicative of high standards, much drive and high self-reliance. A low score describes an individual with low standards, little motivation for achievement and considerable dependence on others.

Factor F measures conscientiousness, orderliness, and acceptance of responsibility. A high score implies the presence of these. A low score indicates the contrasting grouping.

In general, if we interpret 3.5 as a mid-point score, it would seem that these therapists are neither emotionally restrained nor unrestrained, as indicated by Factor A scores, and are somewhat more friendly than hostile with authority figures, as indicated by Factor B scores. At the same time, on Factor C they are rated as having considerably less than average tension, instability, and anxiety. Factor D indicates a somewhat average sense of personal adequacy. On Factors E and F, therapists show considerably greater than average motivation for long term goals and also much conscientiousness and concern for order.

This description of the typical therapist is relative to the norms incorporated into the various scales. It would be of interest to compare these averages to a presumed normal population of similar age, education, and socio-economic status. Although normal population ratings are not available, comparative data are available on a sample of 176 neurotic patients. A comparison of these ratings with a pooled score of the two peer ratings shows some striking differences as given in Table 2. It will be noted that all "*t*" values except one are significant beyond the .001 level, with the greatest differences occurring in Factors B, C, D, and E. It would seem that the patients are rated as much more anxiously tense, with much more hostile rebelliousness, lower sense of personal adequacy, and with much less motivation for long-term goals than the therapists who work with these patients.

The reliability of these therapists' ratings is of obvious pertinence. It is also pertinent to note that inter-judge agreement, involving complex clinical judgments,

TABLE 2. DIFFERENCES BETWEEN MEAN FACTOR SCORES OF THERAPISTS AND PATIENTS

Factor	Therapist	Patient	t
A	3.57	3.75	1.98*
B	3.24	4.07	11.23**
C	2.66	4.09	14.20**
D	3.38	4.18	10.00**
E	4.10	3.42	8.82**
F	4.32	3.99	4.20**

** significant beyond the .001 level
* significant beyond the .05 level

is often considerably lower and more variable than other measures of reliability. Hunt[2] has pointed out some of the problems involved in obtaining a suitable measure of reliability in such instances.

As one approach to the question of reliability we obtained product-moment correlations between self-ratings and an average of the two peer ratings for each of the 25 items in the rating scale. This pooling was done in an attempt to obtain a more stable measure of peer ratings to compare with self-rating. Using a z transformation[1] we obtained an average correlation of .66 for the resulting 67 correlations. The range of these correlations is fairly wide, with three correlations just below .20. However, 55 of the 67 correlations range from .53 to .87 ($\sigma_r = .204$ when $r = 0$) which, for inter-judge agreements is in keeping with findings such as in Hunt[2].

These latter correlations might have been somewhat higher if it were not for the restricted range of the ratings. These were all six point scales and the range was even more restricted by virtue of the fact that the scales had originally been devised for rating neuropsychiatric patients. The patients would have more extreme personality characteristics than the therapists and the range of scores for this latter group would consequently be narrower. As an indication of this difference it should be noted that the standard deviations of the factor scores in Table 1 average about .5 of a rating interval. The comparable standard deviations of ratings for the patient sample in Table 2 are just about twice as large.

At the same time, there is another and opposing influence which may tend to raise these correlations. To what extent are the ratings providing a personality description with general applicability? In other words, is there a basic stereotype which partially describes all the therapists? Since these are all persons in a trio of highly related professions it may well be that individuals in these professions have considerable personality similarities. Furthermore, as was pointed out above, these scales were designed to describe a patient population. To the extent that a sample from another population did not demonstrate any of the extreme characteristics of the patient population there would be some agreement of rating. As one approach to this question we took 15 self ratings at random, 5 psychiatrists, 5 social workers, and 5 psychologists each and obtained 15 product-moment correlations by comparing each member of each group of 5 with the other two groups respectively. The correlations were all positive, with a range from .25 to .68, with 5 of the 15 coefficients above .50. Again using a z transformation we obtained an average correlation of .48 for the 15 correlations, which is greater than twice the standard error for a single such correlation. Our data would thus seem to support the conclusion that there is some general similarity among all the persons rated. Or to be more precise, these therapists tend to describe themselves similarly.

While this latter finding should be taken into account in any interpretation of a coefficient of reliability, it does not necessarily decrease the stability or meaningfulness of the mean scores given in Tables 1 and 2. Considering the nature of the data, it seems quite likely that the mean factor scores are fairly stable and the difference between patient scores and therapist scores indicates a significant difference between the two groups. Furthermore, the very similar ratings, both self and peer,

of psychotherapists in the three disciplines have implications for the team approach which would warrant further exploration.

Summary

Self and peer personality ratings on the MSRPP were obtained on a group of 67 psychotherapists in VA mental hygiene clinics. These included psychiatrists, psychologists, and social workers. Analysis in terms of six previously identified factors showed no significant differences among the three disciplines on any of the factors. At the same time, there were significant differences between these therapists and a sample of patients who had also been rated on all the factors. Therapists were rated as less tense and anxious than the patients, with less hostile rebelliousness, a higher sense of personal adequacy, and more motivation for long term goals.

References

1. Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd Ed.) New York: McGraw-Hill, 1950.
2. Hunt, W. A., Arnhoff, F. N. and Cotton, J. W. Reliability, change, and fantasy in interjudge agreement among clinicians. *J. clin. Psychol.*, 1954, *10*, 294-296.
3. Lorr, M. and Rubinstein, E. A. Personality patterns of neurotic adults in psychotherapy. *J. consult. Psychol.*, 1956, *20*, 257-263.
4. Luborsky, L. The personality of the psychotherapist. *Menninger Quart.*, 1952, *6*(*4*), 1-6.

THE SCATTER PATTERN OF A SOUTHERN NEGRO GROUP ON THE WECHSLER-BELLEVUE INTELLIGENCE SCALE

JULIAN C. DAVIS

Florida State Hospital, Chattahoochee, Florida

Introduction

When so called comparable groups of whites and Negroes are tested, psychometric results of the Negro group almost invariably are inferior.[2, 4, 11] In our own experience it has been found that over half of all Negroes examined at the Florida State Hospital scored within the defective range on standard individual intelligence tests. Despite the group being partially composed of mental patients, the high incidence of deficiency is obviously a psychometric artifact as substantiated through work histories and clinical observation of the individuals. This is not unusual when it is considered that the intelligence scales in current use were standardized on essentially non-Negro populations, a circumstance which leaves us in the lamentable predicament of having no instrument available for the accurate assessment of the Southern Negro. The need for such an instrument has led to the present preliminary investigation.

Even though the position has been challenged,[8, 9] most investigators are in agreement that "an individual of a given socio-economic status has a particular cultural experience, training, and motivation, which have developed certain areas of his mental behavior and certain skills more than others".[3] Through studying a sufficient number of Wechsler protocols of Southern Negroes we hope to learn which areas and skills have developed most. Not only will this information make possible a more accurate interpretation of the findings, which Wechsler[1] says need special

[1]Wechsler, D. Personal communication January, 1949.

interpretation, but should furnish valuable information concerning a choice of problems that are equally common and equally motivating for this culture group in the construction of a new test.

The basic philosophy underlying this study is stated in the observation of Finner[2] that "We are possibly unscientific in attempting to measure the Negro by a yardstick originally constructed for whites. Eventually it seems that we will need to develop a scale that will permit the unique qualities of the Negro to manifest themselves if there are such unique qualities. For the time being it will not matter where the beginning is made. The need is for some objective work in the field."

Procedure

The control group for this study is composed of twenty-seven Negro hospital employees, both men and women. All were given the Full Scale, Wechsler-Bellevue, Form II. The age range was from 18 to 60 years with the median being 22 years. Nine subjects were under twenty, and six over thirty, with only one being over forty.

The patient sample is composed of thirty-three Negro mental patients, men and women, chosen at random from the patient files of the Florida State Hospital irrespective of diagnosis. The only selective factor was that the Full Scale of the Wechsler-Bellevue, Form II had to be completed. Age range of the patient group was 14 to 59 with the median age of 35. Four patients were under twenty and six over forty.

The education of both groups ranged from no school years completed through one year of college. The number of years of school completed for the study groups compares favorably with that of nonwhites 25 years old and over in the State of Florida, as of 1950 (12). Table 1 gives the comparative figures.

Table 1. Comparative Years of School Completed By Test Groups and Nonwhite Persons 25 Years Old and Over for the State of Florida, 1950.

Groups	Years of school completed							Median school years completed
		Elem. school		High school		College		
	None	1-4	5-8	1-3	4	1-3	4 or more	
Nonwhite Population	8%	34%	40%	10%	5%	2%	2%	5.8
Patients (N = 33)	9%	36%	42%	6%	3%	3%		5.2
Employees (N = 27)	7%	33%	33%	15%	7%	4%		6.2

Results

The IQ's of the control group ranged from a low of 51, earned by a twenty-one year old man with a third grade education, to a high of 99, achieved by a thirty-three year old man with one year of college. (Incidentally, the man with the latter score is obviously of superior intelligence despite his "average" IQ. He served as a staff sergeant in the Army and was an excellent attendant on the receiving service at the State Hospital.) The mean IQ for the group is 68. The IQ range of the patient group is from a low of 45 earned by a man with no formal schooling to a high of 108 by a man with one year of college. The mean IQ for the group is 67.

More frequently than not verbal ability exceeds performance ability. There is a wider variability of sub-test scores on the verbal scale, and at the higher levels performance scores are sometimes above the verbal scores. However, discrepancies large enough to be significant are usually found to be in favor of verbal ability. The test scatter follows the same general pattern in both groups. This pattern can be seen most clearly by an examination of the group mean weighted scores in Table 2.

[2]Finner, P. F. Personal communication January, 1949.

TABLE 2. MEAN WEIGHTED SCORES OF SOUTHERN NEGRO GROUPS ON THE WECHSLER-BELLEVUE FORM II.

Sub-tests	Employees (N = 27)	Patients (N = 33)
Information	4	5
Comprehension	7	6
Digit Span	4	4
Arithmetic	5	5
Similarities	3	4
Vocabulary	8	6
Picture Arrangement	5	3
Picture Completion	7	6
Block Design	4	5
Object Assembly	6	7
Digit Symbol	3	3

DISCUSSION

Digit Span and Digit Symbol, along with Similarities, proved most difficult for our sample. Digit Span correlates very poorly with all other tests of general intelligence[10] and Digit Symbol handicaps individuals not practiced in the use of paper and pencil. Wechsler[10] found Digit Span to correlate best with Digit Symbol and least with Comprehension, a test upon which the sample did well. The implication is that Southern Negro scores are depressed most by tests that are not measures of general intelligence for this group.

The low score on Similarities shows that the group has difficulty in differentiating between essential and superficial likenesses and suggests immaturity of thought. Environmental factors might well account for this lack of maturity and no attempt is made to conclude that race *per se*, (at least as represented by the Southern Negro) is a limiting factor in intelligence. In fact, additional supportive evidence of the influence of environmental factors can be drawn from the higher scores obtained on Comprehension and Vocabulary than on Information and Arithmetic. The latter scores are influenced to a greater degree by schooling.

SUMMARY AND CONCLUSIONS

This is a preliminary report of the performance of two Southern Negro groups, 33 mental patients and 27 hospital employees on the Wechsler-Bellevue Form II. The sample is too small to admit definite conclusions but the results suggest a distinct sub-test scatter pattern for the Southern Negro with a probable relationship between environmental factors and sub-test difficulty. There is the added implication that Southern Negro scores may be depressed most by tests which are not measures of general intelligence for this group.

REFERENCES

1. ALTUS, W. D. A note on group differences in intelligence and the type of test employed. *J. consult. Psychol.*, 1948, 3, 194-195.
2. BULLOCK, H. A. A comparison of the academic achievement of white and Negro high school graduates. *J. educ. Res.*, 1950, 44, 179-192.
3. DAVIS, W. A. and HAVIGHURST, R. J. The measurement of mental symptoms. *Sci. Mon.*, 1948, 66, 301-316.
4. HAMMER, E. F. Comparison of the performance of Negro children and adolescents on two tests of intelligence, one an emergency scale. *J. genet. Psychol.*, 1954, 84, 85-93.
5. JENKINS, M. D. The upper limit of ability among American Negroes. *Sci. Mon.*, 1948, 5, 399-401.
6. JENKINS, MARTIN D. and RANDALL, C. M. Differential characteristics of superior and unselected Negro college students. *J. soc. Psychol.*, 1948, 27, 187-202.
7. LEWIS, H. Innovations and trends in the contemporary Southern Negro. *J. soc. Issues*, 1954, 10 (1), 19-27.
8. MCGURK, F. On white and Negro test performance and socio-economic factors. *J. abnorm. soc. Psychol.*, 1953, 48, 448-450.
9. MCGURK, F. Socio-economic status and culturally-weighted test scores on Negro subjects. *J. appl. Psychol.*, 1953, 37, 276-277.
10. WECHSLER, D. The measurement of adult intelligence, third edition. Baltimore: Williams & Wilkins, 1944.
11. YOUNG, F. M. and COLLINGS, J. J. Results of testing Negro contact syphilitics with W-B intelligence scale. *J. soc. Psychol.*, 1954, 39, 93-95.
12. Detailed characteristics Florida, 1950 population census report P-c10. Washington: U. S. Gov't Print. Off., 1952.

THE UTILIZATION OF A PSYCHOLOGIST IN WARD ADMINISTRATION[1]

WARREN W. WEBB AND HIRAM L. GORDON
Vanderbilt University *V. A. Hospital, Fort Lyon, Colorado*

Introduction

Traditional professional roles become firmly established with the passage of time and then are resistant to modification [3, 5]. Under unique and pressing local conditions a reappraisal of job requirements and professional skills can bring about changes in professional roles [1, 4]. Such was the case locally. There being a desire on the part of the Director of Professional Services[2] to provide knowledgeable and interested patient planning despite a growing shortage of psychiatrists, the psychology service was asked to assign a psychologist to play a leading role as a member of a ward team under the general supervision of an interested psychiatrist. After the initial resistance by psychologists and psychiatrists had been worked through, an assignment was made. It was a two year assignment and was to be reviewed upon completion of an experimental evaluation. This paper reports the results of that evaluation.

The goal of the project was the broad but usual one of structuring a total situation within which the individual patient could move in the direction of greater hospital freedom, return to the community, and remission. The elements of such a structured situation which were strived for (although not always attained) were (a) closer cooperation and fuller communication among the supervising psychiatrist, psychologist, social worker, nurses, aides, physical medicine therapists, and special services personnel; (b) increased staff-patient interaction through interviews, individual and group psychotherapy, and coordinated work and recreational activities; (c) increased contact with families including a family discussion group organized by the social worker, and (d) more individualized patient planning. The inevitable interpersonal conflicts under these conditions of shifting roles and work emphases did occur but will not be recounted here since we have no general conclusions as to their resolution. It was our experience that they can only be worked out uniquely by the persons involved.

Procedure and Results

Subjects were 132 patients (E) on the experimental ward and 40 control patients (C) on five other hospital wards. All the patients were male, with a diagnosis of schizophrenia, and were without known organic involvement. Pertinent data for comparison purposes is presented in Table 1.

The variables expected to show the consequences of increased therapeutic effort were as follows: Pre- and post-morbidity scores on the Multidimensional Scale for Rating Psychiatric Patients [2], mean proportion of project time on ground privileges, mean proportion of project time on leave of absence, and mean proportion of project time out of the hospital (trial visit and discharge status). These results are

Table 1. Pre-Project Data

Factors	E (N = 132)		C (N = 40)	
	Mean	SD	Mean	SD
Age	32.8	5.4	33.7	5.4
Education	9.0	3.4	9.4	2.8
Years since first hospitalization	6.6	3.1	6.7	3.2
Morbidity score	32.2	11.5	33.3	11.0

[1]This study was carried out while both authors were on the staff of V. A. Hospital, Roanoke, Virginia. The authors wish to express their gratitude to Drs. C. W. Grady, H. P. Morgan, J. A. Newcome, G. I. Meyer, B. M. Smith, B. J. Fitzgerald, Miss Dorothy McGriff, and to a host of others too numerous to name for making this study possible.

[2]Dr. H. P. Morgan, now Manager, V. A. Hospital, Fort Lyon, Colorado.

shown in Table 2. (For administrative reasons E group lost ten patients to other wards and C group lost one.)

TABLE 2. RESULTS OF PROJECT

Variable	E (N = 122)	C (N = 39)	t	P
Mean morbidity score decrease	2.5	1.2	—	—
Mean proportion of project time				
On privileges	.431	.421	—	—
On leave of absence	.012	.008	—	—
Out of hospital	.127	.029	2.18	< .05

All the measures show trends in the expected direction with time out of the hospital attaining statistical significance. Since the mean morbidity scores[3] do not show a significant drop for E group as compared with C group it cannot be said that basic psychotic processes were generally alleviated by the therapeutic program. Nor can it be said from the privileges and leaves of absence data that patient freedom was substantially increased during the project period. The time out of the hospital data, however, reveal noteworthy success of the planning program in aiding patients to move back into the community.

Two fortuitous happenings should be mentioned. The tranquilizing drugs came into widespread use after the inception of this project and could not be experimentally controlled. Of E group patients, 17 percent received reserpine and 27 percent received chlorpromazine. With C group patients the percentages were 15 and 67, respectively. The results of the project, then, cannot be attributed to the use of tranquilizing drugs. The second happening was an outbreak of tuberculosis in the E group ward during which 21 patients (16 percent) were transferred to TB hospitals and hence were no longer able to benefit from the therapeutic program. Furthermore, the functioning of the entire therapeutic program was hampered by this outbreak.

SUMMARY AND CONCLUSIONS

The professional skills of a psychologist were called upon in a non-traditional assignment to play a leading role in ward administration under the general supervision of a psychiatrist. The total effort of the psychologist and the other members of the ward team was compared with that of other ward teams (without a psychologist in an administrative capacity) in terms of the usual goal of assisting patients toward greater hospital freedom, return to the community, and remission. The variables expected to show the consequences of increased therapeutic effort (ratings of morbidity, time on ground privileges, time on leaves of absence, and time out of the hospital) showed trends in favor of the special project ward. On one of these variables (time out of the hospital) the difference between the project and control wards was statistically significant. It was concluded that the professional skills of a psychologist can be of definite value in coordinating the total efforts of ward team members in patient planning so as to move the chronic schizophrenic patient back toward the community.

REFERENCES

1. GIEDT, F. H. and SANDERS, R. Selection of chronic patients with discharge potential. *V. A. program Guide*, 1956, G-8, M-2, Part X, 39-44.
2. LORR, M. Multidimensional scale for rating psychiatric patients. *V. A. tech. Bull.*, 1953, TB 10-507.
3. RAIMEY, V. C. (Ed.) *Training in clinical psychology*. New York: Prentice-Hall, 1950.
4. SCHLOSSER, J. R., BUEHLER, R. E. and SANDERS, R. Pilot study for the treatment of chronic patients. *V. A. info. Bull.*, 1954, IB 10-68, 3-10.
5. SHAKOW, D. Psychology and psychiatry: A dialogue. *Amer. J. Ortho-psychiat.*, 1949, 19, 191-208, 381-396.

[3]Since the raters may have had at least incidental acquaintance with the patients on the project ward, the possibility of bias was considered. A new staff psychologist, who had no information as to the identity of the patients, interviewed and rated 10 E patients and 10 C patients. Comparison of his ratings with those of the regular raters failed to reveal any trend toward bias.

EVALUATION OF MODIFIED ADMINISTRATION OF THE CALIFORNIA PSYCHOLOGICAL INVENTORY

LAWRENCE A. BENNETT AND ALVIN RUDOFF

Deuel Vocational Institution, Tracy, California

PROBLEM

There is a growing interest and use of the California Psychological Inventory as both a diagnostic and research tool. Since a representative sample is desired in standardization projects, many subjects with limited reading ability are included. This focuses attention on a problem of the effect of limited reading ability on the results obtained with any paper-and-pencil test attempting to assess personality.

It is assumed that the C. P. I. requires a reading ability of about sixth grade level since it is very similar in item content to the M. M. P. I. whose readability has been estimated at this level.[7] In many situations we attempt to assess individual personality changes of a group that includes many who fall below this level. We assume on the basis of numerous previous studies and accepted educational theory that the speaking or aural recognition vocabulary is substantially higher than the reading vocabulary of the group. In this event, it would follow that reading the items would make for better, more accurate and reliable responses. However, experience in the preliminary administration of the C.P. I. in this manner to undisciplined groups (*e.g.*, juvenile delinquents), indicated the possibility of adverse influence on individual responses of spontaneous group reactions such as laughter or side comments. The question then becomes, are the adverse effects of group reaction sufficient to overcome the advantages that would obtain as the result of reading the test to the group? Our concern, then, is to evaluate the effect of the modification of administration on the results; are the results consistent with those obtained from the regular administration or, if they vary, to what extent?

METHOD

A group of inmates at Deuel Vocational Institution, Tracy, California, was used as subjects. A reading ability of at least seventh grade, as measured by the California Achievement Test, was required as the only criterion of selection. This insured that all subjects would be able to read the items fairly adequately. The first 200 inmates so selected during the regular guidance processing were used in the sample. Half of the group read the test items and then, within a three day period, had the items read to them. In the case of the other half, the procedure was reversed.

RESULTS

With the sequence effects controlled by the design, the results from the two groups, called "Reading" and "Read To", were correlated. On each of the subtests the correlations were substantially greater than test-retest reliability coefficients reported by Gough[6] with the single exception of the "Fl" scale (Flexibility). The low correlation might well be related to the small number of items (twenty-two), for when the means of the two groups are compared statistically, no significant difference is found (see Table 1).

However, most clinicians are more interested in individual profiles than in group tendencies. To approach this problem—the objective analysis of a complete profile—the coefficient of profile similarity was applied.[1] This included modifications to handle segments of a profile of zero slope[3] and an index independent of the order of the subscales.[5] Since the C. P. I. comprises 18 subscales, a total of 153 slope comparisons were made on each subject. Of the 200 profiles compared, 95.5% were found to be statistically similar at the 1% level of confidence. Only 3% fell outside the 5% level of confidence for significant similarity and could thus be considered as being not similar.

TABLE 1. COMPARISON OF COEFFICIENTS OF CORRELATION BETWEEN CURRENT TEST-RETEST AND RELIABILITY MEASURES REPORTED BY GOUGH[6]

Scale	D. V. I. Inmates (N = 200)	High School Juniors (N = 101)
Re	.85	.65
To	.87	.71
Fl	.49	.60*
St	.80	.62
Do	.80	.64
Sp	.84	.68
Fe	.73	.59
De	.80	.65
Ie	.80	.74
Ac	.79	.60
Hr	.71	.63
Py	.53	.48
Sr	.80	.60
Im	.86	.75
Sa	.71	.67
In	.58	.38
Gi	.81	.69
Ds	.75	.71
*Fl Scale	Reading	Read-To
Raw Score Mean	8.575	8.515
Sigma	3.41	3.41
Difference Between Means	.06	
t	.25 (not significant)	

SUMMARY AND CONCLUSIONS

Half of a group of 200 subjects were administered the California Psychological Inventory in the usual manner of having the subject read the items to himself. Then, within a three day period, the inventory was administered by reading the items aloud to the subjects. The other half of the group went through the two forms of the administration, but in the reverse order. Inasmuch as the test-retest coefficients of correlation using the two methods of administration compared favorably with the test-retest measures of reliability using the same administration throughout, it is concluded that, as far as groups are concerned, no appreciable difference in result occurs due to the modification of administration.

Furthermore, as the profiles resulting from the two methods are statistically similar in 97% of the cases, it can be concluded that modification of administration has but an insignificant effect on individual profiles. Thus, where low reading ability is a problem, the modified administrative procedure of reading the items to the subjects can be employed without seriously distorting either group trends or individual profiles. This modified procedure has application not only in research studies, but also in clinical screening.

BIBLIOGRAPHY

1. DU MAS, F. M. A quick method of analyzing the similarity of profiles. *J. clin. Psychol.*, 1946, 2, 80-83.
2. DU MAS, F. M. The coefficient of profile similarity. *J. clin. Psychol.*, 1949, 5, 123-131.
3. DU MAS, F. M. A note on the coefficient of profile similarity. *J. clin. Psychol.*, 1950, 6, 300-301.
4. DU MAS, F. M. Quick methods for the analysis of the shape, elevation, and scatter of profiles. *J. clin. Psychol.*, 1953, 9, 345-348.
5. DU TOIT, J. M. A modification of the Index of Profile Similarity. *J. clin. Psychol.*, 1954, 10, 384-385.
6. GOUGH, HARRISON G. A Preliminary Guide for the Use and Interpretation of the California Psychological Inventory, prepared for private distribution, May 10, 1954, Institute of Personality Assessment and Research, University of California, Berkeley, 4, California.
7. JOHNSON AND BOND. The Reading Ease of Commonly Used Tests. *J. appl. Psychol.*, 1950, 34, 319-324.

SEASONAL CHANGES IN ADMISSION RATES OF A STATE HOSPITAL[1]

PAUL A. HAUCK

East Moline State Hospital, Illinois

Introduction

Adverse economic, social and political conditions can have a direct effect upon mental status, and as a result, psychiatric referrals to clinics and admission rates to mental hospitals seem to follow a pattern with such regularity as to make future admission and referral rates somewhat predictable. However, other patterns seem to be constant even in times when the economic, social, or political conditions are in a state of relative rest. There seem to exist seasonal fluctuations in the rate of patient admissions which have been commonly observed, but as yet only impressionistic evidence indicates which seasons seem to favor high or low admission rates to state hospitals. Work in this area has dealt mainly with noting the incidence of mental disease in particular states, or trends as they appear over a period of years, but the question of whether a consistent pattern emerges from the monthly admission rate to state hospitals has remained largely unexplored. The present study undertakes to quantify the admission rates for the East Moline State Hospital on a monthly basis for the years 1953, 1954, and 1955. The population was broken down into sex, and 12 diagnostic groups as recommended in the 1952 *Diagnostic and Statistical Manual on Mental Disorders.*

Results

The total admissions for 1953 to 1955 are presented in Table 1. In 1953 the highest admission rate was in November with 63 and the lowest in October with 46. December takes a sharp drop after the November rise which is then followed by a sharp rise during January. In February, the rate again drops. A careful perusal of each following year reveals that this basic relationship between these five months holds. Taken as a group the six months from September to February contribute about an equal share of the years' admissions, when compared to the remaining six months. But among themselves a pattern appears which remains year after year and which holds also when the three years are totaled up. Within this six month period, small fluctuations from year to year cause October and December to change position for the low-rate month, and November and January to change position for the high-rate month. Aside from these slight changes it appears that the stability of this pattern is being determined by other than chance factors.

In order to determine if this monthly pattern was purely a local phenomenon for the East Moline State Hospital the admission rates of three other state hospitals for 1955 from widely differing areas and social settings were grouped together. These hospitals are at Peoria, Elgin, and Alton, all in Illinois. The total admission rates by months for the four hospitals combined are found in Table 1.

Once again the pattern observed for the East Moline State Hospital is basically repeated when all four hospitals are grouped as one. A slight drop from September to October makes October the low month for the year. This is again followed by a rise in November. Then a drop to a second low in December after which the rates soar to their highest point in January. February follows with its usual dip.

To determine if the large difference between December and January is mainly a function of the alcoholic populations which one could conceivably expect after the holiday season, the entire alcoholic population was subtracted from the rates for the four hospitals. This manipulation causes some minor changes, September is now the lowest month and October is a close third. June has one more admission than Jan-

[1]The author is grateful to Phillip W. Wenig, Supervisor of the Statistical Research Section for the State of Illinois for making the data for this study available to him.

TABLE 1. TOTAL ADMISSIONS TO EAST MOLINE STATE HOSPITAL 1953-1955, FOUR HOSPITALS FOR 1955, WITH AND WITHOUT ALCOHOLICS

Hospitals	Year	Months											
		July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June
East Moline	1953	61	51	59	46	63	51	58	47	53	56	61	55
	1954	49	69	58	43	54	42	50	44	45	51	64	65
	1955	54	56	53	40	59	39	65	54	49	55	58	58
	Total	164	176	170	129	176	132	173	145	147	162	183	178
Four Hospitals	1955	451	472	394	380	428	386	481	418	444	454	432	456
Four Hospitals Minus Alcoholics	1955	362	375	304	309	332	306	384	344	374	362	341	385

uary, making it the high month. Again, however, the pattern shown above is still intact. October is low followed by a November rise. December is low followed by a January rise and the usual February dip.

Summary and Conclusions

The months in question each have a characteristic "social" definition. December, for example, because it is the month during which Christmas falls is regarded by our culture as a time of festivity and good-will. Preparation for December 25 begins soon after Thanksgiving and a mood of elation and anticipation is maintained until January 1. There would seem then to be good reason why December has a low admission rate year after year.

But, what of the entire five month period? We note that October is low. It is the month of beautiful colors, bountiful harvests, and a return to more intimate family life in the cool evenings. The children are sent to school with great relief by the harassed mothers and they are still allowed to play outdoors after school. The hectic summer is over and old routines are revived. When November appears, the cold blasts of winter with its occasional treacherous snows force the children to play more indoors than out. The nights are long and many families are confined to close living quarters which now require large and expensive amounts of fuel for warmth. And admission rates rise.

In January the dead of winter is to be coped with. There is a let down when the holiday wreaths and decorations are packed away. Long periods of cold increase the fuel and clothing bills, and the prospect of paying for an abundant Christmas is sometimes anxiously received. Colds, coughs and other "winter" sicknesses and the expensive medicines to cure them seem to pile up during these bleak days. License plates and taxes probably increase the financial burden.

These problems are probably brought to solution in February by those who do not come to the hospital in January. They have the prospect of looking for the coming of Spring within a month. The days are getting longer and the cold is occasionally broken by warmer spells.

March itself is a paradox. At times the rates drop from the February level, and at times they go up. No explanation is offered at this time for its erratic behavior or for that of the other months.

THE HOWARD INK BLOT TEST

A NEW PROJECTIVE TECHNIQUE

- **Derived experimentally by James W. Howard, Ph.D.**
- **An individual test consisting of 12 plates 8½" x 11". The blots are somewhat larger than the Rorschach and with a wider range of stimulating features. Blots of varied colors selected from an experimental sample of 10,000 blots.**
- **Not intended to be a parallel test to the Rorschach. Statistical findings from a "normal" group show the test to differ markedly from the Rorschach in ways which should indicate greater sensitivity.**

Price: $12.50 (In U. S. and Canada; foreign postage $.50 extra)

HOWARD INK BLOT TEST MANUAL

MONOGRAPH SUPPLEMENT NO. 10 FROM JULY 1953 ISSUE

Price $2.00

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

PERSONALITY CHANGES FOLLOWING TRANSORBITAL LOBOTOMY[1]

NANCY B. ALGEO[2] AND MAXWELL S. PULLEN[3]

Ohio University

INTRODUCTION

Robinson, Freeman and Watts[1, 2] have developed two instruments for the measurement of personality changes in lobotomy patients: the Sensibility Questionnaire (SQ) and the Self-regarding Span test (SRS). The first purports to measure "self-adjustability" or "susceptibility to social pressures;" the second is said to measure "self-continuity". Robinson administered these instruments to 68 lobotomy cases after they had undergone psychosurgery. As controls she used ". . . 12 individuals of similar age and diagnosis who improved without operation." The two groups were also equated on the basis of Stanford-Binet Vocabulary scores. Robinson, *et al.*, found significantly lower scores in the operated group than in the control group. They concluded that lobotomy has the effect of reducing both "susceptibility to social pressure" and "self-continuity", as these terms were defined operationally [1, 2]. This experiment did not provide for an actual preoperative-postoperative control group. They compared the postoperative scores of the lobotomized patients to the scores of the control group that did not receive lobotomy, but it appears that the experimental and control groups were not necessarily comparable. The very fact that the controls did not receive psychosurgery suggests that their personalities may have differed from those of the experimental group prior to the latters' operation. These differences, if there were any, could have been in the direction of less "self-continuity" and "susceptibility", which were the very changes that were alleged to have occurred as a *result* of lobotomy. While this situation is not one which negates the value of the quoted study, it would seem well worth while to note the results when preoperative status of subsequent lobotomy patients is considered. This study, then, is an attempt to repeat the main part of the experiment by Robinson, Freeman and Watts, with modifications to effect the necessary controls.

METHOD

Subjects. The subjects were 34 mental patients at the Athens (Ohio) State Hospital, who had been designated to undergo transorbital lobotomy by Dr. Walter Freeman. Of this group, 13 were men, 21 women. The age range was 25 to 73 with a mean of 38.9. Mean age of the men was 37.3, of the women 39.9. The following diagnostic categories were represented: 11 paranoid schizophrenic, 9 catatonic schizophrenic, 4 manic-depressive depressed, 3 undifferentiated schizophrenic, 2 hebephrenic schizophrenic, 2 paranoid state, 2 involutional disorder, and 1 psychoneurotic disorder. Length of education varied from no formal schooling to high school graduation, with a mean of 9.1 grades completed. IQ's (mainly Wechsler-Bellevue, Form I) ranged from 65 to 116 with the majority falling in the dull-normal and normal categories. Duration of hospitalization ranged from two months to eight and one-half years. Mean length of hospitalization was 22.9 months. Onset of the mental disorder was gradual in 29 cases and acute in the remaining five.

Tests. The Robinson-Freeman Self-regarding Span test (SRS)[1] was used to measure "self-continuity". The term "self-regarding span" is thus defined operationally as the number of seconds that an individual can spend thinking and talking about himself as a person. To measure "susceptibility" the Robinson-Freeman Sensibility Questionnaire (SQ)[1] was used. This test is designed to measure the

[1]The authors are indebted to the administration and staff of the Athens State Hospital, whose cooperation and assistance made this study possible.

[2]Now at Washington, Pennsylvania.

[3]This study is based on a portion of the first author's MA thesis. The second author, who directed the thesis, is responsible for the preparation of this report.

individual's susceptibility to emotional experience in two areas: his concern for his own past and future, and his concern for the opinions of other people.

Procedure. The 34 subjects were divided randomly into two groups, designated I and II. An empirical check, after the division was effected, revealed that the groups were virtually equivalent with respect to the control variables of sex, age, education, diagnostic category, duration of illness, and type of onset. Group I was given the SRS preoperatively and the SQ postoperatively; Group II was given the SQ preoperatively and the SRS postoperatively. The pretests were administered from two to four days prior to the operation and the posttests from three to six days after the operation. The test-retest interval ranged from six to eight days, and the means of the intervals of the two groups did not differ significantly. Pre-postoperative differences were computed for the SRS (Group I preoperative scores minus Group II postoperative scores), and for the SQ (Group II preoperative scores minus Group I postoperative scores). In other words, the total group acted as its own preoperative-postoperative control. This procedure was based on the assumption of comparability referred to above. The differences between group means were tested for significance by means of the standard *t* test. Preoperative-postoperative differences were also computed separately for males and females, and for long *vs.* short hospitalization categories.

Results

The reduction in mean scores on the SQ was 5.47 (from 6.18 to 0.71), a difference significant at the .05 level of confidence. The mean score reduction on the SRS was 4.49 (from 7.12 to 2.63), a difference significant at the .01 level. Thus, while our results are not as marked as Robinson's (owing probably to the smaller number of cases), the present study does confirm her conclusion: that lobotomy tends to reduce "susceptibility" and self-continuity scores.

Two subsidiary analyses of the data yielded the following results: (1) The pre-post differences on both tests were significant for the males as a group (.01 level); for the females as a group neither instrument yielded a significant change. (2) With regard to length-of-hospitalization there was a significant (.05 level) reduction in scores on the SRS for those hospitalized less than 1½ years; this group did not show a significant reduction on the SQ. For the sub-group hospitalized over 1½ years neither test yielded a significant change.

Summary

In an effort to repeat Robinson and Freeman's experiment under more controlled conditions, the SRS Test and the SQ were administered to a group of lobotomy patients before and after the operation. Significant reduction in both self-adjustability and self-continuity were found to occur following lobotomy, thus confirming the evidence obtained by Robinson and Freeman.

References

1. Robinson, M. F. and Freeman, W. *Psychosurgery and the self.* New York: Grune and Stratton, 1954.
2. Robinson, M. F., Freeman, W. and Watts, J. W. Personality changes after psychosurgery. *Amer. Psychologist,* 1949, 4, 272. (Abstract).

EDITORIAL OPINION

SELF-CONSISTENCY IN SCIENTIFICALLY-ORIENTED CLINICAL PRACTICE

After listening to lectures dealing with the validity of many techniques currently used in clinical psychology, many students report feelings of bewilderment and frustration upon discovering that many popular techniques are either unvalidated or actually invalid. We often have had students comment: "You have not left us much to work with. Is it not better to use unvalidated methods than to be completely scientific and have to admit that we don't know anything at all about many problems?" Such a dilemma occurs, for example, with reference to psychoanalytic interpretations of projective test results. In some installations, clinical psychologists routinely are expected to formulate interpretive reports which go far beyond anything which is scientifically valid. Some of these reports soar to speculative heights which would do credit to necromancers. The situation can become pretty embarrassing when such reports are analysed sentence by sentence in order to discover just what of proven validity actually exists to justify such interpretations. The student who attempts to be scientifically critical may find himself in a very delicate position if he attempts to question the validity of what he is being taught. Indeed, considerable prestige may fallaciously accrue to cult members who may achieve transient success by virtue of their ingenuity in parroting popular credos.

There is involved here the extremely important issue of whether the clinician is to rigorously adhere to scientific principles at whatever the cost at all times or only when it is convenient. There can be no compromise with thoroughgoing scientific criticality because any lapse of scientific methodology tends to invalidate the whole process. An apt analogy exists in the sterile techniques of modern surgery, *i.e.*, the whole objective is lost if sterile technique is broken at any point. The clinician is in a particularly difficult position because methods and interpretations of proven validity are so scarce in relation to the tremendous demands placed upon us by society. The best which most clinicians can do is to clearly indicate the point at which exact science leaves off and clinical intuition takes up. The young clinician must not be ashamed to say that he doesn't know. At least we can gain the negative merit of pointing out what is not true even though we can't as yet make positive statements as to what is valid.

We believe that this is the issue over which may experimentalists reach a parting of the ways with their clinical brethren. The psychologist with rigorous experimental-statistical training is perhaps one of the most self-critical people alive because his training causes him to scrutinize every conclusion critically and with reflex automaticity. Nothing is left unsupported or unchallenged. Is it any wonder that the the experimentalist expresses his perplexity and tends to reject with righteous indignation many of the methods and conclusions reached by clinicians in their everyday practice? Let us spend no time rationalizing many of these practices because they are scientifically indefensible. This writer, for example, has been unable to utilize many projective methods and psychoanalytic interpretations because earlier experimental-statistical training has made it simply impossible to assume a sort of split personality, adhering to rigorous scientificality in some areas and becoming uncritical in applying unvalidated techniques in the clinical field. The need to be self-consistent within the broad framework of experimental-statistical science must impose very stringent limitations which we have an obligation to live up to. Let us not be clinically-split personalities, scientific up to a certain point only.

Lest the student be tempted because of his own insecurities to identify himself with some of the more glamorous but unvalidated credos of our time, we hasten to

remind him that some of the most respected figures of clinical and applied psychology have been rigorous scientists whose lasting devotion to basic principles has won them more permanent respect than some meteoric theorists whose status rapidly fades out at the moment of their most apparent brightness. Of the older generation of master clinicians, who achieved higher recognition than Frederick L. Wells, Walter V. Bingham, Lewis M. Terman or Grace H. Kent, and yet who were more modest, unassuming and consistently conservative in not attempting conclusions unsupported by valid data? Their claim for lasting respect must rest upon self-consistent applications of scientific principles. Clinical psychology as science will have its development greatly impeded if we allow ourselves to be distracted by specious dialectics and theorizing by latter day prophets who may be very imaginative but are certainly not scientific.

F. C. T.

BOOK NOTICES

CRONBACH, L. J. AND GLESER, G. C. *Psychological Tests and Personnel Decisions.* Urbana, Ill.: University of Illinois Press, 1957, pp. 165. $3.50

This important book reports research on the relationship of psychological measurements to decision theory. Going beyond the use of tests for the purpose of accurate estimation, the authors examine the relevance of measurements to arrive at qualitative decisions and the loss resulting from erroneous decisions. A realistic appreciation of the difficulties of considering the total contributions of tests against all possible criteria makes us aware of the comparable complexity of factors involved in decision making. Among the topics discussed in separate chapters are types of personnel decisions, characteristics of decision problems, selection decisions with single stage testing, two-stage sequential decisions, placement decisions, efficient testing procedures, the Band-Width-Fidelity problem, classification decisions, evaluation of outcomes and the importance of decision theory. Because the argument is couched largely in terms of complex mathematical models, the reader needs to be adept in higher mathematics to comprehend the full significance of what is presented.

CLARK, KENNETH E. *America's Psychologists: A Survey of a Growing Profession.* Washington, D. C.: American Psychological Association, 1947, pp. 246.

This important study of American psychologists was sponsored by the Policy and Planning Board of the American Psychological Association. Objective evidence is presented concerning the development and present status of psychology, indices of eminence, background and early training of psychologists in various areas of specialization, and the utilization of psychology in the United States.

HALL, CALVIN S. AND LINDZEY, GARDNER. *Theories of Personality.* New York: John Wiley, 1957, pp. 572.

As psychology evolves slowly from metaphysical philosophizing to the status of an exact experimental-statistical science, there is a continuing need for comprehensive reviews in various areas such as perception, learning, motivation and personality. The subject matter of psychological theorizing is so extensive and complex that we

need to have periodic review and evaluation to separate the wheat from the chaff. The authors are to be congratulated on having accomplished a most unbiassed and eclectic evaluation of current theories of personality.

FARNSWORTH, P. R. (Ed.) *Annual Review of Psychology. Volume 8, 1957.* Palo Alto, Calif.: Annual Reviews, Inc., 1957, pp. 502. $7.00

This volume continues in the usual competent tradition of the *Annual Reviews* to present a comprehensive survey of contemporary American psychology. Clinical psychologists will be particularly interested in chapters on individual differences by J. J. Jenkins and D. T. Lykken, personality by C. W. Eriksen, abnormalities of behavior by M. E. Wright, psychotherapy by C. L. Winder, assessment techniques by J. R. Wittenborn and counseling by F. J. Shaw.

ENG, HELGA. *The Psychology of Child and Youth Drawings.* New York: Humanities Press, 1957, pp. 205. $9.50.

This book is a continuation of *The Psychology of Children's Drawings* in which the author reported the development of her niece's drawings through age 8. This volume presents her further development from ages 9 to 24 with evaluative comment and reviews of the literature on the development and psychology of drawing.

REMMERS, H. H., RIVLIN, H. N., RYANS, D. G. AND RYDEN, E. R. *Growth, Teaching and Learning.* New York: Harper, 1957, pp. 557. $4.50

An authoritative book of readings on the growing child, emotional development and mental hygiene, and measurement, evaluation and research techniques in educational psychology.

GROTJAHN, MARTIN. *Beyond Laughter.* New York: McGraw-Hill, 1957, pp. 285. $6.00

Essays on how humor and art speak to and from the unconscious to give man freedom in a culture of repression.

LEVINE, EDNA S. *Youth in a Soundless World.* New York: New York University Press, 1957, pp. 217. $5.00 The psychology of the deaf child.

NORTHCOTT, C. H. *Personnel Management.* New York: Philosophical Library, 1956, pp. 428. $10.00 Third edition.

BALINT, MICHAEL. *The Doctor, the Patient and the Illness.* New York: International Universities Press, 1957, pp. 355. $7.50

ALEXANDER, FRANZ AND STAUB, HUGO. *The Criminal, the Judge and the Public.* Glencoe, Ill.: Free Press, 1957, pp. 239. $4.00 Revised edition.

Child Psychiatry Service

THE MENNINGER CLINIC

THE SOUTHARD SCHOOL

A residential school, for elementary grade children with emotional and behavior problems.

THE CHILDREN'S CLINIC

Outpatient psychiatric and neurologic evaluation of infants and children to eighteen years.

J. Cotter Hirschberg, M. D., Director — Topeka, Kansas; Telephone 3-6494

READY NOW | AN IMPORTANT NEW BOOK

MENTAL DEFICIENCY

IN RELATION TO PROBLEMS OF GENESIS, SOCIAL AND OCCUPATIONAL CONSEQUENCES, UTILIZATION, CONTROL AND PREVENTION

by

J. E. WALLACE WALLIN, Ph.D.

PRICE: $5.00 - - - ORDER NOW

JOURNAL OF CLINICAL PSYCHOLOGY

5 Pearl Street — Brandon, Vermont

THERAPEUTIC SCHOOLS • VOCATIONAL COMMUNITIES • REMEDIAL CAMPS

UNIQUENESS

Some children have unique needs -- needs beyond the usual psychological and psychiatric evaluations. With many children it is found that for every therapeutic experience in the professional office or clinic, there are countless other experiences that are not therapeutic -- in a tense or conflictual home, in a schoolroom that does not meet their needs, or with peers with whom they cannot cope.

It is to serve these children that Devereux Schools have designed various milieu programs in which a plan for each child can be designed so that he learns with teachers of special education while he lives in an environment organized so as not to frustrate or threaten him.

Professional inquiries should be addressed to John M. Barclay, *Director of Development*, or Charles J. Fowler, *Registrar*, Devereux Schools, Devon, Pa. For the western states, address Joseph F. Smith, *Superintendent*, or Keith A. Seaton, *Registrar*, Devereux Schools, Santa Barbara, Calif.

"HUMANI NIHIL A ME ALIENUM PUTO"
THE DEVEREUX FOUNDATION
CALIFORNIA PENNSYLVANIA

Devereux Schools

UNDER THE DEVEREUX FOUNDATION

HELENA T. DEVEREUX, *Director*

Professional Associate Directors

ROBERT L. BRIGDEN, Ph.D. EDWARD L. FRENCH, Ph.D.
MICHAEL B. DUNN, Ph.D. J. CLIFFORD SCOTT, M.D.

Santa Barbara, California **Devon, Pennsylvania**